CONTRIBUTORS TO THIS VOLUME

Zdeněk Kopal
František Link
J. M. Saari
R. W. Shorthill
Donald F. Winter

ADVANCES IN
ASTRONOMY and ASTROPHYSICS

edited by

ZDENĚK KOPAL

Department of Astronomy
University of Manchester
Manchester, England

Volume 9

ACADEMIC PRESS New York and London • 1972

CONTRIBUTORS TO VOLUME 9

ZDENĚK KOPAL, *Department of Astronomy, University of Manchester, England*

FRANTIŠEK LINK,* *Astronomical Institute of the Czechoslovak Academy of Sciences, Prague, Czechoslovakia*

J. M. SAARI,† *Boeing Scientific Research Laboratories, Seattle, Washington*

R. W. SHORTHILL, *Boeing Scientific Research Laboratories, Seattle, Washington*

DONALD F. WINTER,‡ *Boeing Scientific Research Laboratories, Seattle, Washington*

* Present address: Centre National de la Recherche Scientifique, Institut d'Astrophysique, Paris, France.

† Deceased 12 January 1971.

‡ Present address: Department of Oceanography, University of Washington, Seattle, Washington.

CONTENTS OF PREVIOUS VOLUMES

Volume 3

Volume 4

Volume 5

Volume 6

Volume 7

Volume 8

The Roche Model and Its Applications to Close Binary Systems

ZDENĚK KOPAL
Department of Astronomy
University of Manchester
Manchester, England

I

INTRODUCTION

Although the Roche Model of stellar configurations—which will be the principal subject of this chapter—made its first appearance in astronomical literature more than a hundred years ago (cf. Roche, 1849, 1851, 1873), for the first eighty years it led a rather shadowy existence on the periphery of cosmology as a limiting case of a configuration whose total mass is confined to a point at its center. As such, it stood as the extreme opposite of the family of models (Maclaurin or Jacobi) consisting of homogeneous compressible fluids; but since very little was known about the actual internal structure of the stars until at least the first decade of this century, the Roche model continued to stand as a sentinel at one end of the extreme range of configurations which could represent the stars through the time of Poincaré (1854–1912).

Since the second decade of this century we learned—largely through the work of Emden, Eddington, and their contemporaries—that the greater part of the mass in stellar interiors is indeed likely to be condensed very near the center, and that the structure of actual stars will come much closer to the Roche model than to a homogeneous one. From extensive numerical integrations carried out soon thereafter by Chandrasekhar (1933) we learned that, for stars whose central density may bear to the mean density a ratio of 100 or more (as is likely to be true of most Main-Sequence stars, let alone for giants), the Roche model of a rotating configuration will represent the actual form of the surfaces within an error of less than one per cent—a truly astonishing degree of accuracy attained with so simple analytic means! And even more astonishing because of the speed with which stellar configurations of increasing central condensations approach asymptotically the form defined by the Roche model.

The same asymptotic behavior continues, moreover, to hold good for stellar configurations distorted by the rotation as well as tides in close binary systems. For such stars, too, the surfaces of constant pressure and density, in hydrostatic equilibrium, coincide with those of constant potential given by the Roche model; and the boundary of zero density becomes a particular case of a family of surfaces over which the potential arising from all forces acting upon it remains constant.

A theory of the equipotential surfaces for fluid stars of arbitrary structure distorted by rotation or tides has previously been developed by Clairaut, Legendre, and Laplace in ascending series of spherical harmonics (for their review, see, e.g., Kopal, 1960); but with an increasing amount of distortion their convergence becomes very slow, and their complexity such that for terms of higher than second order the entire procedure bogs down. On the other hand, the respective equipotentials for a mass-point model can be written down in a simple algebra form which is *exact* for any double-star configuration consisting of centrally condensed components *irrespective of their proximity or mass-ratio.* No wonder that the quantitative properties of such equipotentials are of great interest for the students of close binary systems, and, in particular, if the latter become eclipsing variables; for, in such a case, many of their geometrical as well as physical properties can be deduced from observations of their light changes and confronted with a theory of far-reaching help for their interpretation.

The aim of the present chapter will be to give a comprehensive outline of such a theory based on work—some of it quite recent—that has so far been done to elucidate the extent of assistance which astronomy of close binary systems can derive from the utilization of the Roche model. Section II of our survey chapter, which follows this Introduction, will be concerned primarily with the geometrical properties of the Roche equipotentials, and, in particular,

of the limiting equipotentials ("Roche limit") at which the respective configuration will cease to be able to contain its mass in a closed volume. As a whole class of stars (the "contact" components of contact or semidetached systems) are known to have attained this limit (see, e.g., Kopal, 1959), geometrical properties of the "Roche limits" are of great importance, and indispensable for a correct interpretation of the light changes of such eclipsing systems.

In Section III of this chapter we shall break new ground for a treatment of many problems concerning close binary systems, by an introduction of a system of "Roche curvilinear coordinates," in which the Roche equipotential surfaces play the role of the radial coordinate ξ, while the angular variables η and ζ are orthogonal to the equipotentials. By virtue of the simplicity with which the distorted surface of a star is defined in such coordinates, their use can facilitate many tasks encountered in the theory of the light- or radial-velocity changes of close eclipsing systems.

Another task eminently suitable for treatment in the new Roche coordinates is the problem of vibrational stability of the Roche model, which will be discussed in Section IV of this chapter. Although definite criteria for stability depend on the structure of the evanescent opaque atmosphere surrounding the central mass-point, the partial differential equations governing small oscillations of such structures prove to be simpler in the Roche coordinates than in any other three-dimensional system; moreover, they permit us to study oscillations around a distorted Roche equipotential rather than those around a sphere (or any other arbitrary figure). Although analytic solutions of such equations turn out to be possible only in the immediate proximity of the origin, their numerical continuation to the (closed) surface of any distortion should offer no difficulty to automatic computing machines.

II

ROCHE EQUIPOTENTIALS

A. *Fundamental Relations*

In order to introduce the Roche equipotentials, let m, m' denote the masses of the two components of a close binary system and R their mutual separation. Suppose, moreover, that the positions of the two stars are referred to a rectangular system of Cartesian coordinates—with the origin at the center of gravity of mass m—the x-axis of which coincides with the line joining the centers of the two stars, while the z-axis is perpendicular to the plane of the orbit. If so, the coordinates of the center of gravity of the system as a whole evidently are

$$\frac{m'R}{m + m'}, 0, 0;$$

and the total potential Ω of combined forces acting at an arbitrary point $P(x, y, z)$ becomes expressible as

$$\Omega = G\frac{m}{r} + G\frac{m'}{r'} + \frac{\omega^2}{2}\left[\left(x - \frac{m'R}{m + m'}\right)^2 + y^2\right], \tag{2.1}$$

where

$$\begin{aligned} r^2 &= x^2 + y^2 + z^2 \\ r'^2 &= (R - x)^2 + y^2 + z^2 \end{aligned} \tag{2.2}$$

represent squares of the distance of P from the centers of gravity of the two components, and ω denotes the angular velocity of rotation of the system about an axis perpendicular to the xy-plane and passing through the center of gravity of the system; lastly, G denotes the constant of gravitation. The first term on the right-hand side of Eq. (2.1) represents the potential arising from the mass of the component of mass m, the second, the disturbing potential of its companion of mass m', and the third, the potential arising from centrifugal force.

Let, furthermore, the angular velocity ω on the right-hand side of Eq. (2.1) be identified with the Keplerian angular velocity

$$\omega^2 = G\frac{m + m'}{R^3}, \tag{2.3}$$

a representation which in close binary systems becomes almost inescapable. If we insert (2.3) in (2.1) and, moreover, adopt m as our unit of mass and R as the unit of length, while the unit of time is chosen so that $G = 1$, Eq. (2.1) may be expressed in terms of spherical polar coordinates

$$\begin{aligned} x &= r\cos\phi\sin\theta = r\lambda, \\ y &= r\sin\phi\sin\theta = r\mu, \\ z &= r\cos\theta \qquad\;\; = r\nu, \end{aligned} \tag{2.4}$$

as

$$\xi = \frac{1}{r} + q\left(\frac{1}{(1 - 2\lambda r + r^2)^{1/2}} - \lambda r\right) + \frac{q+1}{2}r^2(1 - \nu^2), \tag{2.5}$$

where

$$\xi \equiv \frac{R\Omega}{Gm} - \frac{m'^2}{2m(m + m')} \tag{2.6}$$

and

$$q \equiv m'/m \tag{2.7}$$

are nondimensional parameters.

The surfaces generated by setting $\xi =$ constant on the left-hand side of Eq. (2.5) will hereafter be referred to as the *Roche Equipotentials*—in honor of Edouard Albert Roche, a distinguished French mathematician (1820–1883), in whose writings such equipotentials made their first appearance.[1] The form of such equipotentials depends evidently on the value of ξ. If ξ is large, the corresponding equipotentials will consist of two separate ovals (see Fig. 1)
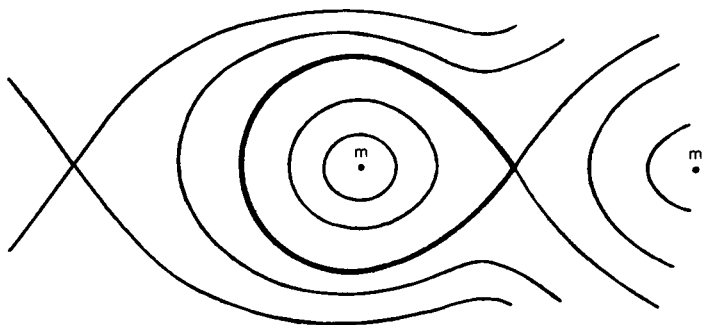
FIG. 1. Geometry (two-dimensional) of the Roche surfaces (schematic). The Roche limit is marked by a heavy line.

closed around each of the two mass-points; for the right-hand side of (2.5) can be large only if r [or $r' = (1 - 2\lambda r + r^2)^{1/2}$] becomes small; and if the left-hand side of (2.5) is to be constant, so must be (very nearly) r or r'. Large values of ξ correspond, therefore, to equipotentials differing but little from spheres—the less so, the greater ξ becomes. With diminishing value of ξ the ovals defined by (2.5) become increasingly elongated in the direction of the center of gravity of the system until, for a certain critical value of ξ_1 characteristic of each mass-ratio, both ovals will unite in a single point on the x-axis to form a dumbbell-like configuration (cf. again Fig. 1) which we propose to call the *Roche Limit*. For still smaller values of ξ the connecting part of the dumbbell opens up and the corresponding equipotential surfaces would envelop both bodies. This latter case is, however, of no direct interest to us in this connection; as for $\xi < \xi_1$ the two initially distinct bodies would coalesce in one and we should no longer have the right to speak of a binary system. In what follows we shall, therefore, limit ourselves to a study of the geometry of surfaces characterized by $\xi \geqslant \xi_1$.

B. *Radius and Volume*

Equation (2.5) of the Roche equipotentials represents an implicit equation defining, for given values of ξ and q, r as a function of λ and ν. When it has

[1] For Roche's life and work, see his obituary by F. Tisserand (1883).

been rationalized and cleared of fractions, the result is an algebraic equation of *eighth* degree in r, whose analytical solution presents unsurmountable difficulties. In the case of pure rotational distortion (obtaining if $q = 0$), Eq. (2.5) can be reduced to a cubic solvable in terms of circular functions. In the case of a pure tidal distortion ($\omega = 0$) Eq. (2.5) becomes a quartic, which could also be solved for r in a closed form (though its solution would be very much more involved). In the general case of rotational *and* tidal distortion interacting, however, any attempt at an *exact* solution of (2.5) for r becomes virtually hopeless, and approximate solutions must inevitably be sought.

In order to obtain them, let us begin by expanding the radical

$$(1 - 2\lambda r + r^2)^{-1/2}$$

on the right-hand side of (2.5) in terms of the Legendre polynomials $P_j(\lambda)$. Doing so and removing fractions we find it possible to replace (2.5) by

$$(\xi - q)r = 1 + q \sum_{j=2}^{\infty} r^{j+1} P_j(\lambda) + nr^3(1 - \nu^2), \tag{2.8}$$

where we have abbreviated

$$n = (q + 1)/2. \tag{2.9}$$

If r is small in comparison with unity (i.e., if the linear dimensions of the equipotential surfaces are small in comparison with our unit of length R), the second and third terms on the right-hand side of (2.8) may be neglected in comparison with unity, in which case, to a first approximation,

$$r_0 = 1/(\xi - q). \tag{2.10}$$

This result asserts that if ξ is large, the corresponding Roche equipotential will differ but little from a sphere of radius r_0.

Suppose now that

$$r_1 = r_0 + \Delta' r = r_0 \left(1 + \frac{\Delta' r}{r_0}\right) \tag{2.11}$$

should represent our next approximation to r. Inserting it in (2.8) we find that

$$1 + \frac{\Delta' r}{r_0} = 1 + q \sum_{j=2}^{\infty} r_0^{j+1} P_j(\lambda) + nr_0^3(1 - \nu^2), \tag{2.12}$$

where, in small terms on the right-hand side, r was legitimately replaced by r_0. The foregoing equation then yields

$$\frac{\Delta' r}{r_0} = q \sum_{j=2}^{4} r_0^{j+1} P_j(\lambda) + nr_0^3(1 - \nu^2) \tag{2.13}$$

correctly to quantities of the order of r_0^5 (i.e., as far as squares and higher terms of first-order distortion remain negligible).

In order to improve upon this approximation let us set, successively,

$$r_2 = r_1 + \Delta'' r = r_0\left(1 + \frac{\Delta' r}{r_0} + \frac{\Delta'' r}{r_0}\right), \tag{2.14}$$

$$r_3 = r_2 + \Delta''' r = r_0\left(1 + \frac{\Delta' r}{r_0} + \frac{\Delta'' r}{r_0} + \frac{\Delta''' r}{r_0}\right), \tag{2.15}$$

$$\vdots \qquad\qquad \vdots$$

$$r_{j+1} = r_j + \Delta^{(j+1)} r = r_0\left(1 + \sum_{i=0}^{j} \frac{\Delta^{(i+1)} r}{r_0}\right), \tag{2.16}$$

where

$$\frac{\Delta^{(i+1)} r}{r_0} = q \sum_{k=3}^{3(N-i)} (r_i{}^k - r_{i-1}^k)\, P_{k-1}(\lambda) + n(r_i{}^3 - r_{i-1}^3)(1 - \nu^2), \tag{2.17}$$

$3N$ denoting the highest power of r_0 to which Eq. (2.16) represents a correct solution for r. We may note that, in general, the leading terms of the expression (2.17) for $(\Delta^{(i+1)} r)/r_0$ will be of $3(i+1)$st degree in r_0; and, similarly, the difference $r_i{}^k - r_{j-1}^k$ in higher terms on the right-hand side of (2.12) will be of the order of r_0^{3i+k}.

Suppose that, in what follows, we wish to construct the explicit form of an approximate solution of Eq. (2.8) in the form of (2.15), correctly to 15 (say) quantities of the order of $(\Delta''' r)/r_0$, which should therefore differ from the exact solution of (2.8) at most in quantities of the order of r_0^{12}. By use of the expression already established for $(\Delta' r)/r_0$ the explicit forms of $(\Delta'' r)/r_0$ and $(\Delta''' r)/r_0$ can successively be found,[2] and their insertion in (2.15) leads to the equation

$$\begin{aligned}\frac{r - r_0}{r_0} = {} & r_0{}^3\{qP_2 + n(1 - \nu^2)\} \\ & + r_0{}^4\{qP_3\} \\ & + r_0{}^5\{qP_4\} \\ & + r_0{}^6\{qP_5 + 3[qP_2 + n(1 - \nu^2)]^2\} \\ & + r_0{}^7\{qP_6 + 7q[qP_2 + n(1 - \nu^2)]\, P_3\} \\ & + r_0{}^8\{qP_7 + 8q[qP_2 + n(1 - \nu^2)]\, P_4 + 4q^2 P_3{}^2\}\end{aligned}$$

[2] For fuller details of this process, see Kopal (1954b, 1959).

$$+ r_0^9\{qP_8 + 9q[qP_2 + n(1 - \nu^2)]\, P_5 + 9q^2P_3P_4$$
$$+ 6[qP_2 + n(1 - \nu^2)]^3 + 6[q^3P_2^3 + n^3(1 - \nu^2)^3]\}$$
$$+ r_0^{10}\{qP_9 + 10q[qP_2 + n(1 - \nu^2)]\, P_6 + 5q^2[P_4^2 + 2P_3P_5]$$
$$+ 45q[qP_2 + n(1 - \nu^2)]^2\, P_3\}$$
$$+ r_0^{11}\{qP_{10} + 11q[qP_2 + n(1 - \nu^2)]\, P_7 + 11q^2[P_3P_6 + P_4P_5]$$
$$+ 55q[qP_2 + n(1 - \nu^2)]^2\, P_4$$
$$+ 55q^2[qP_2 + n(1 - \nu^2)]\, P_3^2\} + \cdots, \tag{2.18}$$

where we have abbreviated $P_j \equiv P_j(\lambda)$, and which represents the desired approximate solution of Eq. (2.8) for r as a function of λ and ν in the form of an expansion in ascending powers of r_0 (as defined by Eq. 2.17).

The volume V of a configuration whose radius vector r is given by the foregoing equation (2.18) will be specified by

$$V = \frac{2}{3}\int_{-1}^{1}\int_{-\sqrt{1-\lambda^2}}^{\sqrt{1-\lambda^2}} \frac{r^3\, d\lambda\, d\nu}{\mu}, \tag{2.19}$$

where $\mu^2 = 1 - \lambda^2 - \nu^2$. By virtue of the algebraic identity

$$r^3 = r_0^3\left(1 + \frac{r - r_0}{r_0}\right)^3 \tag{2.20}$$

we find it convenient to express the integrand in (2.19) in terms of (2.18) as a function of λ and ν. This integrand will, in general, consist of a series of terms of the form $\lambda^m\nu^n/\mu$, factored by constant coefficients; therefore, the entire volume V will be given by an appropriate sum of partial expressions $V_n{}^m$ of the form

$$V_n^m = \int_{-1}^{1}\int_{-\sqrt{1-\lambda^2}}^{\sqrt{1-\lambda^2}} \frac{\lambda^m\nu^n}{\mu}\, d\lambda\, d\nu. \tag{2.21}$$

These expressions vanish (on grounds of symmetry) if either m or n is an odd integer. If, however, both happen to be even and such that $m = 2a$ and $n = 2b$, an evaluation of the foregoing integrals readily reveals that

$$V_{2i}^{2a} = \frac{\sqrt{\pi}\, \Gamma(a + \frac{1}{2})\, \Gamma(b + \frac{1}{2})}{\Gamma(a + b + \frac{3}{2})}, \tag{2.22}$$

where Γ denotes the ordinary gamma functions. As, accordingly,

$$\int_{-1}^{1}\int_{-\sqrt{1-\lambda^2}}^{\sqrt{1-\lambda^2}} \frac{P_j(\lambda)\, d\lambda\, d\nu}{\sqrt{1 - \lambda^2 - \nu^2}} = \begin{cases} 2\pi & \text{if } j = 0, \\ 0 & \text{if } j > 0, \end{cases} \tag{2.23}$$

and

$$\int_{-1}^{1}\int_{-\sqrt{1-\lambda^2}}^{\sqrt{1-\lambda^2}} \frac{\nu^{2j}\,d\lambda\,d\nu}{\sqrt{1-\lambda^2-\nu^2}} = \frac{2\pi}{j+1}, \tag{2.24}$$

we eventually find that the volume of a configuration whose surface is a Roche equipotential will be given by

$$\begin{aligned} V = \tfrac{4}{3}\pi r_0^3(1 &+ \tfrac{12}{5}q^2r_0^6 + \tfrac{15}{7}q^2r_0^8 + \tfrac{18}{9}q^2r_0^{10} + \cdots \\ &+ \tfrac{22}{7}q^3r_0^9 + \tfrac{157}{7}q^3r_0^{11} + \cdots \\ &+ 2nr_0^3 + \tfrac{32}{5}n^2r_0^6 + \tfrac{176}{7}n^3r_0^9 + \cdots \\ &+ \tfrac{8}{5}nqr_0^6 + \tfrac{296}{35}nq(2q+n)\,r_0^9 \\ &+ \tfrac{26}{35}nq(q+3n)\,r_0^{11} + \cdots), \end{aligned} \tag{2.25}$$

correctly to quantities of the order up to and including r_0^{11}. With n and r_0 as given by Eqs. (2.9) and (2.10) the volume V becomes an explicit function of ξ and q alone and can be tabulated in terms of these parameters.

C. *Roche Limit*

It was pointed out already in Section II,A that a diminution of the value of the constant ξ on the left-hand side of Eq. (2.5) will cause the respective Roche equipotentials to expand from nearly spherical configurations to ovals of increased elongation in the direction of the attracting center until, for a certain critical value of ξ characteristic of each mass-ratio, these ovals unite in a single point on the line joining their centers. Such configurations represent the largest *closed* equipotentials capable of containing the whole mass of the respective components, and will hereafter be referred to as their *Roche limits*. Any star filling its Roche limit will, moreover, be termed a *contact component*; and a binary system consisting of a pair of such components, a *contact system*. The fact that close binaries in which one, or both, components have attained their Roche limits actually exist in considerable numbers (see Kopal, 1954a, 1956, 1959) adds importance to a study of the geometry of Roche limits in binary systems of different mass-ratios.

In order to do so, our first task should be to specify the values of ξ for which the two loops of the critical equipotential (cf. Fig. 2) develop a common point of contact at P_1 ; but its determination presupposes a knowledge of the position
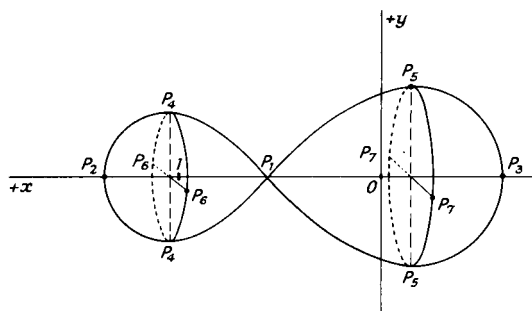
FIG. 2. Three-dimensional geometry (schematic) of a contact binary at the Roche limit.

of P_1 on the x-axis. The location of this point is characterized by the vanishing of the gravity due to all forces, which means that, at that point,

$$\xi_x = \xi_y = 0. \tag{2.26}$$

Now a differentiation of (2.5), rewritten in terms of rectangular coordinates, with respect to x and y yields

$$\xi_x = -xr^{-3} + q[(1 - x)(r')^{-3} - 1] + 2nx, \tag{2.27}$$

$$\xi_y = -y[r^{-3} + q(r')^{-3} - 2n], \tag{2.28}$$

$$\xi_z = -zr^{-3} - qz(r')^{-3}, \tag{2.29}$$

where $r^2 = x^2 + y^2 + z^2$ and $r'^2 = (1 - x)^2 + y^2 + z^2$ continue to be given by Eqs. (2.1) and $2n = q + 1$ in accordance with (2.9).

The partial derivative ξ_y evidently vanishes everywhere along the x-axis; but the vanishing of the former renders the x-coordinate of P_1 to be a root of the equation

$$x^{-2} - x = q[(1 - x)^{-2} - (1 - x)], \tag{2.30}$$

which, after removal of the fractions, assumes the form

$$(1 + q)\,x^5 - (2 + 3q)\,x^4 + (1 + 3q)\,x^3 - x^2 + 2x - 1 = 0. \tag{2.31}$$

For $q = 0$ the foregoing equation would evidently reduce to

$$(1 - x)^3\,(x^2 + x + 1) = 0, \tag{2.32}$$

the value $x = 1$ becoming a triple root. Therefore, for small values of q, the root x_1 of Eq. (2.30) which is interior to the interval $0 < x < 1$ should be approximated by the expansion

$$x_1 = 1 - w + \tfrac{1}{3}w^2 + \tfrac{1}{9}w^3 + \cdots \tag{2.33}$$

in terms of the auxiliary parameter

$$w^3 = q/3(1 + q); \tag{2.34}$$

and more accurate values of x_1 can further be obtained by the method of differential corrections.

Once a sufficiently accurate value of x_1 has thus been obtained, the actual value of ξ corresponding to our critical equipotential follows as

$$\xi_1 \equiv \xi(x_1, 0, 0). \tag{2.35}$$

Moreover, the points $P_{4,5}$ in the xy-plane (see again Fig. 2) are evidently characterized by the vanishing of the derivative dy/dx at the Roche limit. Their coordinates $x_{4,5}$ and $y_{4,5}$ can, therefore, be evaluated by solving the simultaneous system

$$\begin{aligned} \xi(x, y, 0) &= \xi_1, \\ \xi_x(x, y, 0) &= 0; \end{aligned} \tag{2.36}$$

and once the values of $x_{4,5}$ have thus been found, the z-coordinates of points $P_{6,7}$ in the xz-plane (cf. Fig. 2) follow as roots of a single equation

$$\xi(x_{4,5}, 0, z) = \xi_1. \tag{2.37}$$

TABLE I[a]

q	x_1	ξ_1	x_4	$\pm y_4$	x_5	$\pm y_5$	$\pm z_6$	$\pm z_7$
1.0	0.50000	3.75000	1.01134	0.37420	−0.01134	0.37420	0.35621	0.35621
0.8	0.52295	3.41697	1.01092	0.35388	−0.01168	0.39501	0.33770	0.37491
0.6	0.55234	3.06344	1.01029	0.32853	−0.01198	0.42244	0.31431	0.39909
0.4	0.59295	2.67810	1.00926	0.29465	−0.01213	0.46189	0.28260	0.43278
0.3	0.62087	2.46622	1.00847	0.27204	−0.01204	0.49015	0.26123	0.45599
0.2	0.65856	2.23273	1.00735	0.24233	−0.01163	0.52989	0.23294	0.48714
0.15	0.68392	2.10309	1.00656	0.22280	−0.01117	0.55774	0.21425	0.50781
0.1	0.71751	1.95910	1.00552	0.19746	−0.01034	0.59609	0.18991	0.53451
0.05	0.76875	1.78886	1.00397	0.15979	−0.00859	0.65804	0.15366	0.57291
0.02	0.82456	1.65702	1.00245	0.11992	−0.00618	0.73070	0.11522	0.61434
0.01	0.85853	1.59911	1.00165	0.09613	−0.00457	0.77779	0.09231	0.62867
0.005	0.88635	1.56256	1.00110	0.07689	−0.00327	0.81807	0.07379	0.64170
0.001	0.93231	1.52148	1.00041	0.04550	−0.00137	0.88816	0.04361	0.65762
0.0002	0.96001	1.50737	1.00015	0.02678	−0.00052	0.93264	0.02566	0.66348
0	1.00000	1.50000	1.00000	0.00000	0.00000	1.00000	0.00000	0.66667

[a] The data collected in this table are taken from Kopal (1959). More extensive tabulations of the same parameters for q = 1.00 (−0.02) 0.10 to 5D have been prepared by Plavec and Kratochvíl (1964). Cf. also ten Bruggencate (1934).

The accompanying Table I lists five-digit values of ξ_1, x_1 ; $x_{4,5}$, $y_{4,5}$; and $z_{6,7}$ for Roche limits appropriate for 15 discrete values of the mass-ratio.

It may further be noticed that if, in place of ξ_1, we introduce a new constant C_1 as defined by the equation

$$C_1 = \frac{2\xi_1}{1+q} + \left(\frac{q}{1+q}\right)^2 = 2(1-\mu)\,\xi_1 + \mu^2, \tag{2.38}$$

where we have abbreviated

$$\mu = \frac{q}{1+q} = \frac{m'}{m+m'}, \tag{2.39}$$

the values of C_1 remain largely invariant with respect to the mass-ratio, and sensibly equal to 4 provided that q does not depart greatly from unity. This is demonstrated by an inspection of the tabulation of C_1 as given in column 2 of Table II. In consequence, the corresponding simple expression

$$\xi_1 = 2(1+q) - \frac{q^2}{2(1+q)} = \frac{4-\mu^2}{2(1-\mu)} \tag{2.40}$$

is found to approximate the exact values of ξ_1 within 1% if $1 \geqslant q \geqslant 0.5$, or within 10% for the wider range $1 \geqslant q > 0.1$.

TABLE II[a]

q	C_1	$(r_0)_1$	$(r_0)_2$	V_1	V_2	$(r^*)_1$	$(r^*)_2$	v_1	v_2
1	4.00000	0.36363	0.36363	0.22704	0.22704	0.37845	0.37845	0.072267	0.072267
0.8	3.99417	0.38212	0.34528	0.26459	0.19374	0.39825	0.35896	0.075799	0.069377
0.6	3.96993	0.40594	0.32199	0.31974	0.15665	0.42420	0.33441	0.081422	0.066485
0.4	3.90749	0.43896	0.29025	0.40923	0.11444	0.46057	0.30115	0.091184	0.063726
0.3	3.84744	0.46163	0.26876	0.48148	0.09089	0.48622	0.27892	0.099619	0.062683
0.2	3.74900	0.49195	0.24018	0.59399	0.06492	0.52147	0.24933	0.113443	0.061996
0.15	3.67456	0.51201	0.22121	0.68002	0.05079	0.54552	0.22973	0.124462	0.061967
0.1	3.57027	0.53789	0.19642	0.80715	0.03564	0.57760	0.20414	0.141308	0.062385
0.05	3.40962	0.57509	0.15931	1.0289	0.01910	0.62626	0.16584	0.17193	0.063854
0.02	3.24945	0.61087	0.11974	1.2700	0.007961	0.67179	0.12387	0.2062	0.06462
0.01	3.16665	0.62928	0.09606	1.4656	0.004042	0.70465	0.09882	0.2356	0.06497
0.005	3.10959	0.64203	0.07686	1.5950	0.002038	0.7248	0.07865	0.2551	0.06520
0.001	3.03992	0.65769	0.04549	1.868	0.0004114	0.764	0.04614	0.298	0.06554
0.0002	3.01414	0.66350	0.02679	2.067	0.0000826	0.790	0.02702	0.329	0.06575
0	3.00000	0.66667	0.00000	2.26663	0.0000000	0.81488	0.00000	0.36075	0.065843

[a] After Kopal (1959).

The mean radii $(r_0)_{1,2}$ of the two components of contact systems become (consistent with Eqs. 2.10 and 2.38) equal to

$$(r_0)_{1,2} = \frac{2(1-\mu)}{C_1 - (1+\mu)^2 + 1}, \tag{2.41}$$

where, for the primary component, $0 \leqslant \mu \leqslant 0.5$; while, for the secondary, $0.5 \leqslant \mu \leqslant 1$. Alternatively, we may fall back on Eq. (2.10) and, by inserting for ξ_1 from (2.35), write

$$(r_0)_1 = \frac{2x_1}{2 + 2qx_1^3(1-x_1)^{-1} + (q+1)\,x_1^3}; \tag{2.42}$$

while $(r_0)_2$ is obtainable from the same expression if we replace x_1 by $1 - x_1$ and q by its reciprocal. The values of $(r_0)_{1,2}$ so determined are listed as functions of the mass-ratio in columns 3 and 4 of Table II. Having evaluated them, we are in a position to invoke Eq. (2.25) for expressing the volumes $V_{1,2}$ of contact components—the reader will find them tabulated in columns 5 and 6 of Table II—while columns 7 and 8 list the equivalent radii $(r^*)_{1,2}$ of spheres having the same volume as the respective contact component. The penultimate and ultimate columns of Table II then contain the quantities

$$v_{1,2} = \frac{\omega^2}{2\pi G\bar{\rho}_{1,2}} = \frac{2}{3}\left(1 + \frac{m_{2,1}}{m_{1,2}}\right)(r^*)^3_{1,2}, \tag{2.43}$$

where ω denotes the (Keplerian) angular velocity of axial rotation of each component and $\bar{\rho}_{1,2}$, their respective mean densities.

The series on the right-hand side of the volume equation (2.25), which are at the basis of our numerical data as given in columns 5–10, converge with satisfactory rapidity if the masses of the two components are not too unequal, but fail to do so if the mass of one component becomes very much larger than the other. In order to attain adequate representation of the radii and volumes in such cases, asymptotic solutions of Eq. (2.8) must be sought as $\mu \rightarrow 0$ or 1.

In order to do so, we find it advantageous to rewrite in the alternative form

$$(1 - 2\lambda r + r^2)\{(1-\nu^2)\,r^3 - 2\lambda\mu r^2 + (\mu^2 - C_1)\,r + 2(1-\mu)\}^2 = 4\mu^2 r^2, \tag{2.44}$$

where C_1 as well as μ are defined by Eqs. (2.38) and (2.39) above. Consider first the case of negligible disturbing mass, when $\mu = 0$. As long as quantities of the order of μ^2 may be ignored, Eq. (2.44) will admit of a real solution only if

$$(1-\nu^2)\,r^3 - 2\lambda\mu r^2 - C_1 r + 2(1-\mu) = 0. \tag{2.45}$$

For small values of μ, the solution of this latter equation can be sought in the form

$$r = S_{10} + S_{11}\mu + \cdots, \tag{2.46}$$

where S_{10}, S_{11},... are defined by the equations

$$(1 - \nu^2) S_{10}^3 - C_1 S_{10} + 2 = 0, \tag{2.47}$$

$$3(1 - \nu^2) S_{10}^2 S_{11} - C_1 S_{11} - 2 = 2\lambda S_{10}^2, \tag{2.48}$$

etc., whose solutions become

$$S_{10} = 2\left(\frac{C_1}{3(1 - \nu^2)}\right)^{1/2} \sin\left[\frac{1}{3}\sin^{-1}\frac{3}{C_1}\left(\frac{3(1 - \nu^2)}{C_1}\right)^{1/2}\right] \tag{2.49}$$

and

$$S_{11} = \frac{2(1 + \lambda S_{10}^2)}{3(1 - \nu^2) S_{10}^2 - C_1}, \tag{2.50}$$

respectively.

Equation (2.46) with its coefficients as given by (2.49) and (2.50) will closely approximate the form of the primary component of a contact system which is very much more massive than the secondary. Its first term S_{10} defines obviously the form of a Roche equipotential distorted by centrifugal force alone. If $\mu \to 0$, $\xi_1 \to 1.5$ and $C_1 \to 3$, in which case the parametric equation of the corresponding critical equipotential assumes the neat form

$$r = \frac{2}{(1 - \nu^2)^{1/2}} (\sin \tfrac{1}{3} \cos^{-1} \nu), \tag{2.51}$$

and its volume V_1, in accordance with Eq. (2.19), becomes

$$\begin{aligned} V_1 &= \frac{32}{3}\pi \int_0^1 (1 - \nu^2)^{-3/2} \sin^3(\tfrac{1}{3}\cos^{-1}\nu)\, d\nu, \\ &= \frac{4}{3}\pi \left\{3\sqrt{3} - 4 + 3\log\frac{3(\sqrt{3} - 1)}{\sqrt{3} + 1}\right\} = 2.26663 \cdots. \end{aligned} \tag{2.52}$$

It is this foregoing value, rather than the one which would follow from a straightforward application of (2.25), which has been used to complete the last entry in column 5 of Table II.

If the primary component accounts thus for most of the total mass of our contact binary system, the volume of the secondary must clearly tend to zero. The form of its surface will, in turn, be given by an asymptotic solution of Eq. (3.44) as $\mu \to 1$. Let us, therefore, expand this solution in a series of the form

$$r = S_{20}(1 - \mu) + S_{21}(1 - \mu)^2 + \cdots; \tag{2.53}$$

inserting it in (3.44) we find the vanishing of the coefficients of equal powers of $(1 - \mu)$ to require that

$$S_{20} = \frac{2}{C_1 - 3}, \qquad S_{21} = -\left(2 + \frac{\lambda}{C_1 - 3}\right) S_{20}^2, \tag{2.54}$$

etc. An application of Eq. (2.19) reveals, moreover, that the volume V_2 of the respective configuration should be approximated by

$$V_2 = \tfrac{4}{3}\pi[(1-\mu)^3 S_{20}^3 - 6(1-\mu)^4 S_{20}^4 - \cdots], \tag{2.55}$$

and the radius $r_2{}^*$ of a sphere of equal volume becomes

$$r_2{}^* = (1-\mu) S_{20} - 2(1-\mu)^2 S_{20}^2 + \cdots. \tag{2.56}$$

A glance at the second column of Table II reveals that as $q \to 0$, $C_1 \to 3$ and, as a result, the product $(1-\mu) S_{20}$ tends to become indeterminate for $\mu = 1$. In order to ascertain its limiting value, let us depart from Eq. (2.38) which, on insertion of ξ_1 from (2.35), assumes the form

$$C_1 = \frac{2(1-\mu)}{1 - x_1} + \frac{2\mu}{x_1} + (1 + \mu + x_1)^2, \tag{2.57}$$

with the root x_1 approximable by means of (2.33) where, by (2.39),

$$3w^2 = 1 - \mu. \tag{2.58}$$

Inserting (2.33) in (2.57) we find that, within the scheme of our approximation,

$$C_1 = 3(1 + 3w^2 - 4w^3 + \cdots), \tag{2.59}$$

so that

$$(1-\mu) S_{20} = \frac{2w}{3 - 4w} + \cdots, \tag{2.60}$$

and therefore

$$r_2{}^* = \tfrac{2}{3}w - \tfrac{23}{27}w^3 + \cdots. \tag{2.61}$$

In consequence, it follows from (2.43) that, for a secondary component of vanishing mass,

$$v_2 = \tfrac{1}{3}(\tfrac{2}{3})^4 (1 - \tfrac{16}{3}w^2 + \cdots). \tag{2.62}$$

An inspection of the last two columns of Table II reveals that, for the primary (more massive) component, the value of v_1 increases monotonously with diminishing mass-ratio m_2/m_1 from 0.07227 for the case of equality of masses to 0.36075 for $m_2 = 0$, at which point the primary component becomes rotationally unstable and matter begins to be shed off along the equator if

axial rotation were any faster. On the other hand, for the secondary (less massive) component the values of v_2 diminish with decreasing mass-ratio from 0.07227 until, as $m_2 \rightarrow 0$, the value of $2^4/3^5$ has been attained.

D. *Geometry of the Eclipses*

The data assembled in the foregoing section on the geometry of contact configuration lead to a number of specific conclusions regarding the eclipse phenomena to be exhibited by such systems. For suppose that a contact binary both components of which are at their Roche limits is viewed by a distant observer, whose line of sight does not deviate greatly from the x-axis of our model as shown in Fig. 2. If so, then in the neighborhood of either conjunction one component is going to eclipse the other, and the system will exhibit a characteristic variation in brightness. If, in turn, the observed light variation is analyzed for the geometrical elements the fractional "radii" $r_{1,2}$ of the two components should (very approximately) be identical with the quantities $y_{4,5}$ as listed in columns 5 and 7 of Table I. In Table III we have, accordingly, listed four-digit values of the sums $r_1 + r_2$ as well as the ratios r_2/r_1 of the "radii" of such contact components as functions of their mass-ratio.

An inspection of this tabulation reveals that, within the scheme of our approximation, *the sum $r_1 + r_2$ of fractional radii of both components in contact binary systems is very nearly constant* and equal to $0.75 \leqslant 0.01$ for a very wide range of the mass-ratios q; whereas the *ratio r_2/r_1 decreases monotonically with diminishing value of q*. Therefore, a photometric determination of the sum $r_1 + r_2$, which,

TABLE III[a]

q	$r_1 + r_2$	r_2/r_1
1	0.7484	1.0000
0.9	0.7486	0.9495
0.8	0.7489	0.8959
0.7	0.7496	0.8389
0.6	0.7510	0.7777
0.5	0.7529	0.7112
0.4	0.7565	0.6379
0.3	0.7622	0.5550
0.2	0.7722	0.4573
0.15	0.7805	0.3995
0.1	0.7935	0.3312
0.05	0.8178	0.2428
0	1.0000	0.0000

[a] After Kopal (1959).

unfortunately, represents nearly all that can be deduced with any accuracy from an analysis of light curves due to shallow partial eclipses, cannot be expected to tell us anything new about contact systems; or, in particular, about their mass-ratios. It is the ratio of the radii r_2/r_1 whose determination would provide a sensitive photometric clue to the mass-ratio of a contact system. This underlines the importance of photometric determination of the ratios of the radii of contact binary systems; but owing to purely geometrical difficulties this important task of light curve analysis is, unfortunately, not yet well in hand.

Suppose next that a contact binary system, consisting of two components at their Roche limits, is viewed by a distant observer from an arbitrary direction. What will be the range of such directions from which this observer will see both bodies mutually eclipse each other during their revolution? In order to answer this question, let us replace the actual form of the corresponding Roche limit by an *osculating cone* which is tangent to it at the point of contact P_1. The equation of this cone may readily be obtained if we expand the function $\xi(x, y, z)$ of Roche equipotentials in a Taylor series, in three variables, about P_2. The first partial derivatives ξ_x, ξ_y, and ξ_z have already been given by Eqs. (2.27)–(2.29) of the preceding section. Differentiating these equations further we find that

$$\xi_{xx} = (3x^2 - r^2)r^{-5} + q\{3(1-x)^2 - r'^2\}(r')^{-5} + q + 1, \tag{2.63}$$

$$\xi_{yy} = (3y^2 - r^2)r^{-5} + q\{3y^2 - r'^2\}(r')^{-5} + q + 1, \tag{2.64}$$

$$\xi_{zz} = (3z^2 - r^2)r^{-5} + q\{3z^2 - r'^2\}(r')^{-5}, \tag{2.65}$$

$$\xi_{xy} = 3xyr^{-5} - 3q(1-x)y(r')^{-5}, \tag{2.66}$$

$$\xi_{xz} = 3xzr^{-5} - 3q(1-x)z(r')^{-5}, \tag{2.67}$$

$$\xi_{yz} = 3yzr^{-5} + 3qyz(r')^{-5}. \tag{2.68}$$

We note that all first (as well as mixed second) derivatives of ξ vanish at P_1. Hence, a requirement that the sum of nonvanishing second-order terms should add up to zero provides us with the desired equation of the osculating cone in the form

$$(x - x_1)^2 (\xi_{xx})_1 + y^2(\xi_{yy})_1 + z^2(\xi_{zz})_1 = 0, \tag{2.69}$$

where

$$\begin{aligned} (\xi_{xx})_1 &= 2p + q + 1, \\ (\xi_{yy})_1 &= -p + q + 1, \\ (\xi_{zz})_1 &= -p, \end{aligned} \tag{2.70}$$

in which we have abbreviated

$$p \equiv x_1^{-3} + q(1 - x_1)^{-3}. \tag{2.71}$$

The direction cosines l, m, n, of a line normal to the surface of this cone clearly are

$$l, m, n = (f_\xi, f_y, f_z)/(f_\xi^2 + f_y^2 + f_z^2)^{1/2}, \tag{2.72}$$

where $f(\xi, y, z)$ stands for the left-hand side of Eq. (2.69) and $\xi \equiv x - x_1$. Morever, the direction cosines of the axis of this cone in the same coordinate system are (1, 0, 0). Consequently, the angle ϵ between any arbitrary line on the surface of the osculating cone and its axis will be defined by the equation

$$\cos(\tfrac{1}{2}\pi - \epsilon) = l, \tag{2.73}$$

or, more explicitly,

$$\tan^2 \epsilon = -\left(\frac{(\xi_{yy})_1 y^2 + (\xi_{zz})_1 z^2}{(\xi_{yy})_1^2 y^2 + (\xi_{zz})_1^2 z^2}\right)(\xi_{xx})_1, \tag{2.74}$$

where the values of ξ_{xx}, ξ_{yy}, and ξ_{zz} at P_1 are given by Eq. (2.70) above.

Suppose now that the orbital xy-plane of the two components is inclined at an angle i to a plane perpendicular to the line of sight (i.e., one tangent to the celestial sphere at the origin of the coordinates), and that ψ_1 denotes the angle of the first contact of the eclipse (as measured from the moment of superior conjunction). If so, then obviously

$$\begin{aligned} \cos\epsilon = x &= \cos\psi_1 \sin i, \\ y &= \sin\psi_1 \sin i, \\ z &= \cos i, \end{aligned} \tag{2.75}$$

in Eq. (2.74), and the latter can be simplified to disclose that

$$\delta^2 = \frac{az^2}{a-1}\left(\frac{a+2-4\delta^2}{a+2-3\delta^2}\right), \tag{2.76}$$

where

$$\delta^2 = y^2 + z^2 = \sin^2\psi_1 \sin^2 i + \cos^2 i \tag{2.77}$$

denotes the apparent projected distance between the centers of both components at the moment of first contact of the eclipse, and

$$a = \frac{q+1}{p} = \frac{q+1}{x_1^{-3} + q(1-x_1)^{-3}}. \tag{2.78}$$

An inspection of Eq. (2.76) reveals several features deserving explicit statement. Thus the *maximum duration of eclipses* (i.e., the maximum value of ψ_1) will

evidently obtain if the line of sight lies in the orbital plane and, therefore, $z = \cos i = 0$; and if so, Eq. (2.76) discloses that

$$\cos^2 \psi_{\max} = \tfrac{1}{3}(1 - a). \tag{2.79}$$

With the aid of the x_1's as given in column 2 of Table I the values of a can be readily evaluated for any mass-ratio q; and the corresponding values of $\psi_{\max}$ are then listed in the second column of Table IV. A glance at these data reveals

TABLE IV[a]

q	$\pm(\psi_1)_{\max}$ (deg)	$i_{\min}$ (deg)
1	57.3122	34.4499
0.8	57.3192	34.4478
0.6	57.3474	34.4388
0.4	57.4280	34.4153
0.2	57.6446	34.3511
0.1	57.9296	34.2690

[a] The data are by Kopal (1959). More extensive tabulation of ψ_1 has since been given by Plavec and Kratochvíl (1964).

that *the values of* $\psi_{\max}$ *are remarkably insensitive to the mass-ratio,* which constitutes a fact of considerable practical importance for the student of close binary systems. The variations of light exhibited by such systems in the course of each cycle are so smooth and continuous that it is often very difficult—if not impossible—to detect by an inspection of the observed light changes just where the eclipses may set in. Our present analysis now supplies a theoretical answer: namely, *no matter what the mass-ratio may be, the light changes of a close binary system are bound to be unaffected by eclipses for all phase angles outside the range of* $\pm 58°$ *even if both components are in actual contact.* Therefore, the light changes exhibited at phases within 65°.2 around each quadrature must be due solely to the distorted form of both components, and may be subject to "rectification" without fear of interference from eclipse phenomena.

The closed form of Eq. (2.76) invites another related question which can be asked in this connection: namely, what is the *minimum inclination* of the orbital plane to the celestial sphere below which no eclipses may occur for contact binary systems? This minimum inclination will evidently be attained if the eclipse becomes grazing just at the time of the conjunctions (i.e., if $\psi_1 = 0$, in which case $\delta = z$). Under these circumstances Eq. (2.76) can be shown to imply that

$$\cos^2 i_{\min} = (2 + a)/(3 - a), \tag{2.80}$$

where a continues to be given by (2.78). A tabulation of the values of $i_{\min}$ as defined by the foregoing equation (2.80) for six discrete values of the mass-ratio can be found in column 3 of Table IV. An inspection of these data reveals that, *regardless of the mass-ratio, no binary system can exhibit eclipses if its orbit is inclined by less than* 34°.4 *to the celestial sphere—even if its components are so close as to be in actual contact.* For any value of i above this limit eclipses may occur (and *must* occur for contact binary systems) of durations requiring the angle ψ_1 of first contact to satisfy the inequality $0° \leqslant \psi_1 \leqslant 58°$. A relation between ψ_1 and i for each particular value of the mass-ratio is then represented by the full-dress equation (2.76); and a compressed tabulation of the values ψ_1 as functions of i and q is given in Table V.

TABLE V[a]

q	$\cos^2 i$							
	0	0.10	0.20	0.30	0.40	0.50	0.60	$\cos^2 i_{\min}$
				(deg)				
1	57.312	54.986	52.180	48.684	44.116	37.673	27.089	0
0.8	57.319	54.995	52.188	48.693	44.128	37.688	27.112	0
0.6	57.347	55.018	52.211	48.714	44.147	37.706	27.129	0
0.4	57.428	55.093	52.282	48.782	44.212	37.775	27.213	0
0.3	57.509	55.167	52.352	48.849	44.277	37.843	27.297	0
0.2	57.645	55.295	52.471	48.963	44.391	37.959	27.437	0
0.1	57.930	55.555	52.716	49.197	44.619	38.195	27.725	0

[a] After Kopal (1959).

E. *External Envelopes*

In Section II,B we were concerned with various properties of Roche equipotentials when $\xi > \xi_1$, and in Section II,D we investigated the geometry of limiting double-star configurations for which $\xi = \xi_1$. The aim of the present section will be to complete our analysis of the geometrical properties of the Roche model by considering what happens when $\xi < \xi_1$. In introductory paragraphs of Section II,A we inferred on general grounds that, if $\xi < \xi_1$, the dumbbell figure which originally surrounded the two components will open up at P_1 (cf. again Fig. 1), and the corresponding equipotentials will enclose *both* bodies.

When will these latter equipotentials containing the total mass of our binary system cease to form a *closed* surface? A quest for the answer takes us back to Eq. (2.27) defining the partial derivative ξ_x. We may note that the right-hand side of this equation is positive when $x \rightarrow \infty$, but becomes negative when

$x = 1 + \epsilon$, where ϵ denotes a small positive quantity. It becomes positive again as $x \rightarrow 0$, and changes sign once more for $x \rightarrow -\infty$. Since ξ_x is finite and continuous everywhere except at $|x| = \infty$ and for $|r = 0|$ or $r' = 0$, it follows that it changes sign *three* times by passing through zero at points x_1, x_2, x_3, whose values are such that

$$\begin{aligned} &\text{(a)} \quad 0 < x_1 < 1, \\ &\text{(b)} \qquad x_2 > 1, \\ &\text{(c)} \qquad x_3 < 0; \end{aligned} \tag{2.81}$$

and of these, only the first one has been evaluated so far, in Section II,C, and its numerical values listed in column 2 of Table I.

An evaluation of the remaining roots $x_{2,3}$ offers, however, no greater difficulty. In embarking upon it we should merely keep in mind that, regardless of the sign of x, the distances r and r' as defined by Eqs. (2.2) are *positive* quantities. Thus, unlike in case (a)—when, by setting $r = x$ and $r' = 1 - x$, we were led to define x_1 as a root of Eq. (2.31)—in case (b), when $x_2 > 1$, we must set $r = x$ but $r' = x - 1$; and in case (c), when $x_3 < 0$, $r = -x$ and $r' = 1 - x$. After doing so and clearing the fractions we may verify that the equation $\xi_x = 0$ in the case of (b) and (c) assumes the explicit form

$$(1 + q)\, x^5 - (2 + 3q)\, x^4 + (1 + 3q)\, x^3 - (1 + 2q)\, x^2 + 2x - 1 = 0 \tag{2.82}$$

and

$$(1 + q)\, x^5 - (2 + 3q)\, x^4 + (1 + 3q)\, x^3 + x^2 - 2x + 1 = 0, \tag{2.83}$$

respectively.

For $q = 0$, the former equation (2.82) becomes identical with (2.2) and reduces to (2.32) admitting of $x = 1$ as a triple root. Hence, for small values of q, the root $x_2 > 1$ of the complete equation (2.82) should be expansible as

$$x_2 = 1 + \left(\frac{\mu}{3}\right)^{1/3} + \frac{1}{3}\left(\frac{\mu}{3}\right)^{2/3} + \frac{1}{9}\left(\frac{\mu}{3}\right) + \cdots \tag{2.84}$$

in terms of fractional powers of $\mu \equiv q/(q + 1)$. Similarly, Eq. (2.83) reduces for $q = 0$ to

$$(x - 1)^2\, (x^3 + 1) = 0, \tag{2.85}$$

admitting of only one negative root: namely, -1. In consequence, the negative root x_3 of (2.83) should, for small values of μ, be approximable in terms of integral powers of μ by an expansion of the form

$$x_3 = -1 + \frac{7}{12}\mu - \frac{1127}{20736}\mu^3 + \cdots. \tag{2.86}$$

The approximate values of x_2 and x_3 as obtained from (2.84) or (2.86) may, moreover, be subsequently refined to any degree of accuracy by differential corrections or any other standard method.

Once sufficiently accurate values of $x_{2,3}$ have thus been established, the values of ξ corresponding to equipotentials which pass through these points can be ascertained from the equation

$$\xi_{2,3} = \xi(x_{2,3}, 0, 0), \tag{2.87}$$

while the corresponding values of $C_{2,3}$ can then be found from

$$C_{2,3} = 2(1 - \mu)\, \xi_{2,3} + \mu^2. \tag{2.88}$$

A tabulation of five-digit values of $x_{2,3}$ and $C_{2,3}$ is given in columns 2–5 Table VI. It may also be noticed that, to a high degree of approximation,

$$\xi_3 \doteq \frac{3}{2} + 2q - \frac{q^2}{2(1+q)} \tag{2.89}$$

or

$$C_3 \doteq 3 + \mu; \tag{2.90}$$

while, somewhat less accurately,

$$(x_2 - 1)^2 = 1 - x_3^2. \tag{2.91}$$

A comparison of the values of $C_{2,3}$ as given in Table VI with those of C_1 from Table II reveals that, for all values of $q > 0$,

$$C_1 > C_2 \geqslant C_3. \tag{2.92}$$

For any value of C within the limits of the inequality $C_1 > C > C_2$ the corresponding equipotential will surround the whole mass of the system by a common *external envelope*, which may enclose the common atmosphere of the two stars. For $C = C_2$, this envelope will develop a conical point P_2 (at which $\xi_x = \xi_y = \xi_z = 0$) at $x = x_2$, i.e., behind the center of gravity of the less massive component (see Figs. 1 or 2); and if $C < C_2$, the respective equipotentials will open up at P_2. For $C = C_3$, a third conical point P_3 develops behind the center of gravity of the more massive component at $x = x_3$; and if $C < C_3$, the equipotentials will become open at both ends. Their intersection with the xy-plane will then no longer represent a single closed curve, but will split up in two separate sections (symmetrical with respect to the x-axis), closing gradually around two points which make equilateral triangles with the centers of mass of the two components. The coordinates of such points are specified by the requirements that $r = r' = 1$; consequently, $x = 0.5$ and $y = \pm\sqrt{3}/2$. These

triangular points represent also the loci at which our equipotentials vanish eventually from the real plane if (consistent with Eqs. 2.5 and 2.38) their constants C reduce to

$$C_{4,5} = 3 - \mu + \mu^2. \tag{2.93}$$

The values of $C_{4,5}$'s as given by this equation are listed in column 6 of Table VI for $1 > q > 0$, and represent the lower limits attainable by these constants; for if $C < C_{4,5}$, the equipotential curves ξ = constant in the xy-plane become imaginary, and thus devoid of any further physical significance.

TABLE VI[a]

q	x_2	C_2	$-x_3$	C_3	$C_{4,5}$
1.0	1.69841	3.45680	0.69841	3.45680	2.75000
0.8	1.66148	3.49368	0.73412	3.41509	2.75309
0.6	1.61304	3.53108	0.77751	3.35791	2.76563
0.4	1.54538	3.55894	0.83180	3.27822	2.79592
0.3	1.49917	3.55965	0.86461	3.22675	2.82249
0.2	1.43808	3.53634	0.90250	3.16506	2.86111
0.15	1.39813	3.50618	0.92372	3.12959	2.88658
0.1	1.34700	3.45153	0.94693	3.09058	2.91735
0.05	1.27320	3.34671	0.97222	3.04755	2.95465
0.02	1.19869	3.22339	0.98854	3.01961	2.98077
0.01	1.15614	3.15344	0.99422	3.00990	2.99020
0.005	1.12294	3.10301	0.99710	3.00498	2.99504
0.001	1.07089	3.03838	0.99942	3.00099	2.99900
0.0002	1.04108	3.01387	0.99988	3.00020	2.99980
0	1.00000	3.00000	1.00000	3.00000	3.00000

[a] The data collected in this table are taken from Kopal (1959). For other tabulations of these quantities—in particular, for very small values of the parameter $\mu = q/(q + 1)$— cf. Rosenthal (1931), Kuiper and Johnson (1956), Szebehely (1967), or Kitamura (1970).

III

ROCHE COORDINATES

In the second part of this chapter we outlined the principal geometrical features characteristic of the Roche equipotentials and, in particular, of the largest closed equipotentials capable of containing their total mass, which we call the "Roche limit." In Section II,B we expressed the external form of such equipotentials, in spherical polar coordinates, by Eq. (2.18) requiring 4, 15, and 45 separate harmonic terms to describe the shape of a Roche equipotential correctly

to terms of the first, second, and third order in r_0, respectively. For (say) $q = 1$, $r_0 = 0.378$ at the Roche limit (cf. Table II), which means that even a retention of 45 terms on the right-hand side of Eq. (2.18) will not permit us to describe the form of the respective equipotential more accurately than within errors of the order of one unit of the fourth decimal place.

On the other hand, an implicit equation

$$\xi(r, \lambda, \nu) \equiv r^{-1} + r'^{-1} - \lambda r + r^2(1 - \nu^2) = 3.75$$

will define the same surface exactly in much simpler algebraic form. Inasmuch as such equipotentials can represent to a high degree of approximation the surface of the components in close binary systems, and the form of such a surface figures prominently in many tasks encountered in double-star astronomy (such as a determination of the light or velocity curves of close eclipsing systems, hydrodynamic studies of the gas streams in such systems, etc.), the idea suggests itself to introduce a curvilinear system of coordinates—hereafter referred to as the *Roche coordinates*—in which one (the ξ-) coordinate is defined by the Roche equipotential surfaces of the form (2.5) given to us in closed algebraic form (and playing thus the generalized role of the radical coordinate in spherical polars), while the angular Roche coordinates η, ζ are defined by the requirements that they be *orthogonal* to ξ as well as with respect to each other.

The first attempt to introduce such coordinates in practical use goes back to Prendergast (1960). This latter investigator considered, however, an orthogonal system only in the xy-plane without determining the shape of the curves $\eta = \text{constant}$; and their generalization to three dimensions resulted, in Prendergast's hands, in a nonorthogonal system. A theory underlying such coordinates in the plane as well in space has not been placed on solid analytical foundations until subsequent work by Kopal (1969, 1970); and their topological properties elucidated on the basis of extensive numerical integrations by Kitamura (1970). In what follows we shall give a brief outline of the analytical theory of such coordinates, and describe their principal properties in space.

The general equations which must be fulfilled by any curvilinear system of orthogonal coordinates are known to be

$$\xi_x\eta_x + \xi_y\eta_y + \xi_z\eta_z = 0, \tag{3.1}$$

$$\xi_x\zeta_x + \xi_y\zeta_y + \xi_z\zeta_z = 0, \tag{3.2}$$

$$\eta_x\zeta_x + \eta_y\zeta_y + \eta_z\zeta_z = 0, \tag{3.3}$$

where the subscripts x, y, z represent partial differention with respect to the respective variable. If so, a transformation of the metric element

$$(dx)^2 + (dy)^2 + (dz)^2 = h_1^2(d\xi)^2 + h_2^2(d\eta)^2 + h_3^2(d\zeta)^2 \tag{3.4}$$

will be specified by the metric coefficients

$$h_1^{-2} = \xi_x{}^2 + \xi_y{}^2 + \xi_z{}^2, \tag{3.5}$$

$$h_2^{-2} = \eta_x{}^2 + \eta_y{}^2 + \eta_z{}^2, \tag{3.6}$$

$$h_3^{-2} = \zeta_x{}^2 + \zeta_y{}^2 + \zeta_z{}^2; \tag{3.7}$$

and, moreoever, the direction cosines of a normal to the surfaces ξ = constant, η = constant, and ζ = constant are given by the ratios

$$l_1 = h_1\xi_x, \quad m_1 = h_1\xi_y, \quad n_1 = h_1\xi_z, \tag{3.8}$$
$$l_2 = h_2\eta_x, \quad m_2 = h_2\eta_y, \quad n_2 = h_2\eta_z, \tag{3.9}$$
$$l_3 = h_3\zeta_x, \quad m_3 = h_3\zeta_y, \quad n_3 = h_3\zeta_z, \tag{3.10}$$

respectively.

The foregoing equations (3.1) to (3.10) hold good for any triply orthogonal set of curvilinear coordinates ξ, η, ζ (cf. Darboux, 1910; Forsyth, 1912). In what follows, we propose to restrict their choice by an identification of their ξ-coordinates with the Roche potential (2.5)—so that surfaces ξ = constant become identical with those of equal potential of centrally condensed stars—and our task will be to specify the form of the corresponding surfaces of constant η and ζ.

A. *Rotational Problem*

In order to approach the general problem raised in the preceding section in steps, consider first the case when $q = 0$ and when the sole reason for the departure of the Roche equipotentials from spherical form is axial rotation with a (normalized) angular velocity, the square of which we shall (consistent with Eq. 2.9) denote by $2n$. If so, the Roche potential (2.5) reduces to

$$\xi(r, \nu) = (1/r) + nr^2(1 - \nu^2), \tag{3.11}$$

where $\nu \equiv \cos\theta$; and the partial first derivatives of ξ with respect to x, y, z, already given by Eqs. (2.27)–(2.29), now reduce to

$$\begin{aligned} \xi_x &= -x(r^{-3} - 2n), \\ \xi_y &= -y(r^{-3} - 2n), \\ \xi_z &= -z(r^{-3}). \end{aligned} \tag{3.12}$$

The lines that are orthogonal to the equipotentials defined by Eq. (3.11) are known to be given by the differential equations

$$\frac{dx}{\xi_x} = \frac{dy}{\xi_y} = \frac{dz}{\xi_z}. \tag{3.13}$$

Since, however, by (3.12)

$$\xi_x/\xi_y = x/y \tag{3.14}$$

regardless of the extent of the distortion, the first part of Eqs. (3.13) readily yields

$$dx/dy = x/y, \tag{3.15}$$

and integrates to

$$x/y = \text{constant} = \cos\eta \text{ (say)}. \tag{3.16}$$

Accordingly, the surfaces $\eta = \text{constant}$ turn out to be identical with the meridional planes $\phi = \text{constant}$ of the rotationally distorted Roche model; and the three-dimensional Roche coordinates in this case prove to constitute a "Lamé family" (cf. Darboux, 1910) of the respective triply orthogonal set.

The properties of the remaining surfaces $\zeta = \text{constant}$ which are orthogonal to ξ and η can then be obtained by an integration of the orthogonality conditions (3.11)–(3.13). Since, however, by (3.16)

$$\begin{aligned}(r\sin\eta)\,\eta_x &= \mu^{-1},\\ (r\sin\eta)\,\eta_y &= -\lambda\mu^{-1},\\ (r\sin\eta)\,\eta_z &= 0,\end{aligned} \tag{3.17}$$

and, in addition, $\zeta_\lambda = 0$ on account of the rotational symmetry of our configuration, Eq. (3.13) happens to be automatically satisfied without imposing any restriction on ζ. Therefore, in order to determine the latter equation (3.12) alone needs to be integrated; and in doing so (cf. Kopal, 1971) we find it to be satisfied by an expression of the form

$$\cos\zeta = \nu \sum_{j=0}^{\infty} (2n)^j\, r^{3j} X_j(\nu), \tag{3.18}$$

where

$$X_0(\nu) = 1, \qquad X_1(\nu) = -\tfrac{1}{3}(1-\nu^2), \tag{3.19}$$

while, for $j > 1$, all subsequent $X_j(\nu)$'s can be generated with the aid of the recursion formula

$$3jX_j + (1-\nu^2)[(\nu X_{j-1})' - 3(j-1)\,X_{j-1}] = 0, \tag{3.20}$$

where primes denote differentiation with respect to ν.

Accordingly, to quantities of second order in n, Eq. (3.18) can be written out as

$$\cos\zeta = \nu[1 - \tfrac{2}{3}nr^3(1-\nu^2) - \tfrac{4}{9}n^2r^6(1-\nu^2) + \cdots], \tag{3.21}$$

and, by (3.16),

$$\eta = \phi. \tag{3.22}$$

Equations (3.11) and (3.22) together with (3.18) constitute the exact form of the triply orthogonal set of Roche coordinates ξ, η, ζ for the rotational problem expressed in terms of the spherical-polar coordinates r, θ, ϕ (or r, λ, ν); with (3.20) representing a second approximation to (3.18). Moreover, by Eqs. (3.5) and (3.6) we find that the metric coefficients

$$h_1(r, \nu) = -\frac{r^2}{[1 - 4nr^3(1 - \nu^2) + n^2r^6(1 - \nu^2)]^{1/2}}, \tag{3.23}$$

$$h_2(r, \nu) = r\sqrt{1 - \nu^2}, \tag{3.24}$$

exactly; while, to a second-order approximation, Eq. (3.7) leads to

$$h_3(r, \nu) = r[1 + \tfrac{4}{3}nr^3(1 - 2\nu^2) + \tfrac{2}{9}n^2r^6(4 - 24\nu^2 + 21\nu^4) + \cdots]. \tag{3.25}$$

Moreover, the direction cosines

$$l_1 = \lambda[1 - 2nr^3\nu^2 - 6n^2r^6\nu^2(1 - \nu^2) + \cdots], \tag{3.26}$$

$$m_1 = \mu[1 - 2nr^3\nu^2 - 6n^2r^6\nu^2(1 - \nu^2) + \cdots], \tag{3.27}$$

$$n_1 = \nu[1 + 2nr^3(1 - \nu^2) + 2n^2r^6(1 - \nu^2)(2 - 3\nu^2) + \cdots], \tag{3.28}$$

$$l_2 = -\mu/\sqrt{1 - \nu^2}, \tag{3.29}$$

$$m_2 = \lambda/\sqrt{1 - \nu^2}, \tag{3.30}$$

$$n_2 = 0, \tag{3.31}$$

and

$$l_3 = (\lambda\nu/\sqrt{1 - \nu^2})[1 + 2nr^3(1 - \nu^2) + 2n^2r^6(1 - \nu^2)(2 - 3\nu^2) + \cdots], \tag{3.32}$$

$$m_3 = (\mu\nu/\sqrt{1 - \nu^2})[1 + 2nr^3(1 - \nu^2) + 2n^2r^6(1 - \nu^2)(2 - 3\nu^2) + \cdots], \tag{3.33}$$

$$n_3 = -\sqrt{1 - \nu^2}\,[1 - 2nr^3\nu^2 - 6n^2r^6\nu^2(1 - \nu^2) + \cdots]. \tag{3.34}$$

The foregoing equations (3.11), (3.18), and (3.22) express the Roche curvilinear coordinates ξ, η, ζ and the associated metric coefficients or direction cosines in terms of the polar coordinates r, θ, and ϕ. On the other hand, from Eq. (2.18) it follows that, in our case,

$$r = r_0[1 + nr_0^3(1 - \nu^2) + 3n^2r_0^6(1 - \nu^2)^2 + \cdots], \tag{3.35}$$

where

$$r_0 = 1/\xi, \tag{3.36}$$

and by an inversion of (3.21) it follows that

$$1 - \nu^2 = \sin^2 \zeta[1 - \tfrac{4}{3}nr_0^3 \cos^2 \zeta - \tfrac{8}{9}n^2r_0^6 \cos^2 \zeta(7 - 8\cos^2 \zeta) + \cdots]. \tag{3.37}$$

it can be shown (cf. Kopal, 1970, 1971) that, correctly to quantities of second order in n,

$$h_1(\xi, \zeta) = -r_0^2[1 + 4nr_0^3 \sin^2 \zeta - \tfrac{4}{3}n^2r_0^6 \sin^2 \zeta(22 - 85 \sin^2 \zeta) + \cdots] \tag{3.38}$$

$$h_2(\xi, \zeta) = r_0 \sin \zeta[1 - \tfrac{1}{3}nr_0^3(2 - 5\sin^2 \zeta) + \tfrac{1}{9}n^2r_0^6(2 - 50 \sin^2 \zeta + 75 \sin^4 \zeta) + \cdots], \tag{3.39}$$

and

$$h_3(\xi, \zeta) = r_0[1 - \tfrac{1}{3}nr_0^3(2 - 7\sin^2 \zeta) + \tfrac{1}{9}n^2r_0^6(2 - 88 \sin^2 \zeta + 145 \sin^4 \zeta) + \cdots], \tag{3.40}$$

while the direction cosines l_j, m_j, n_j, given by (3.26)–(3.34) in terms of spherical polar coordinates, can be rewritten (with the aid of 3.35–3.37) in terms of the Roche coordinates ξ, η, ζ if required with equal ease.

B. *Tidal Problem*

If a configuration built up in accordance with our Roche model is nonrotating (i.e., $n = 0$) but distorted by tides ($q > 0$), the expression (2.8) for its potential will reduce to

$$\xi = \frac{1}{r} + q\left(\frac{1}{(1 - 2\lambda r + r^2)^{1/2}} - \lambda r\right); \tag{3.41}$$

and its derivatives, to

$$\begin{aligned} \xi_x &= -xr^{-3} + q[(1 - x)\, r'^{-3} - 1], \\ \xi_y &= -y(r^{-3} + qr'^{-3}), \\ \xi_z &= -z(r^{-3} + qr'^{-3}); \end{aligned} \tag{3.42}$$

in accordance with (2.27)–(2.29). Since now

$$\xi_x/\xi_z = y/z \tag{3.43}$$

in place of (3.14) of the rotational problem, the second part of Eq. (3.13) now leads to

$$dy/dz = y/z, \tag{3.44}$$

and integrates to

$$y/z = \text{constant}; \tag{3.45}$$

so that only one integral of the system (3.1)–(3.3) needs to be obtained to specify the triply orthogonal system of Roche coordinates associated with tidal distortion, of which (by virtue of 3.45) one surface will again constitute a family of planes.

On performing the requisite integration (for its details, see Kopal, 1970) we find that, correctly to terms of *first* order in superficial distortion,

$$\eta = \cos^{-1}\lambda - \frac{q}{\sqrt{1-\lambda^2}} \sum_{j=2}^{4} \frac{r^{j+1}}{j+1} P_j'(\lambda), \tag{3.46}$$

while, consistent with (3.46)

$$\zeta = \cos^{-1}(\nu/\sqrt{1-\lambda^2}). \tag{3.47}$$

The geometrical meaning of the surfaces defined by constant values of the curvilinear coordinates η and ζ are simple: the equation η = constant represents a family of distorted cones passing through the origin and symmetrical with respect to the x-axis (i.e., one of symmetry of the tidal distortion); while the surfaces ζ = constant are planes passing through the x-axis and perpendicular to the plane $x = 0$.

The metric coefficients $h_{1,2,3}$ associated with the transformation (3.4) in the case of the tidal distortion continue to be given by Eqs. (3.5)–(3.7), in which we insert the requisite partial derivatives of (3.41), (3.46), and (3.47). The results disclose that

$$h_1(r, \lambda, \nu) = r^2 \left(1 + q \sum_{j=2}^{4} j r^{j+1} P_j(\lambda)\right), \tag{3.48}$$

$$h_2(r, \lambda, \nu) = r \left\{1 + q \sum_{j=2}^{4} \frac{r^{j+1}}{j+1} [j(j+1) P_j(\lambda) - \lambda P_j'(\lambda)]\right\}, \tag{3.49}$$

and

$$h_3(r, \lambda, \nu) = r\sqrt{1-\lambda^2}; \tag{3.50}$$

the last of which is exact, but the first two represent first-order approximations.

Moreover, the direction cosines of the normals to the surfaces ξ = constant, η = constant, and ζ = constant follow again from Eqs. (3.8)–(3.10) as[3]

$$l_1 = \lambda - q(1-\lambda^2) \sum_{j=2}^{4} r^{j+1} P_j'(\lambda), \tag{3.51}$$

$$m_1 = \mu \left(1 + q\lambda \sum_{j=2}^{4} r^{j+1} P_j'(\lambda)\right), \tag{3.52}$$

[3] Kopal and Kitamura (1968) have given expressions for l_1, m_1, n_1 correct to terms of second order in small quantities.

$$n_1 = \nu \left(1 + q\lambda \sum_{j=2}^{4} r^{j+1} P_j'(\lambda)\right); \tag{3.53}$$

$$l_2 = \sqrt{1 - \lambda^2} \left(1 + q\lambda \sum_{j=2}^{4} r^{j+1} P_j'(\lambda)\right), \tag{3.54}$$

$$m_2 = -\frac{\mu}{\sqrt{1 - \lambda^2}} \left(\lambda - q \sum_{j=2}^{4} r^{j+1} (1 - \lambda^2)\, P_j'(\lambda)\right), \tag{3.55}$$

$$n_2 = -\frac{\nu}{\sqrt{1 - \lambda^2}} \left(\lambda - q \sum_{j=2}^{4} r^{j+1} (1 - \lambda^2)\, P_j'(\lambda)\right); \tag{3.56}$$

and

$$l_3 = 0, \tag{3.57}$$

$$m_3 = \nu/\sqrt{1 - \lambda^2}, \tag{3.58}$$

$$n_3 = -\mu/\sqrt{1 - \lambda^2}; \tag{3.59}$$

the last three of which are again exact.

On the other hand, if we wish to express these quantities in terms of Roche curvilinear rather than spherical-polar coordinates, this task can be accomplished by noting from (2.18) that, to the first order in small quantities,

$$r(\xi, \eta) = r_0 \left(1 + q \sum_{j=2}^{4} r_0^{j+1} P_j(\lambda_0)\right) \tag{3.60}$$

and

$$1 - \lambda^2 = (1 - \lambda_0^2) \left(1 + 2q \sum_{j=2}^{4} \frac{r_0^{j+1}}{j + 1} \lambda_0 P_j'(\lambda_0)\right), \tag{3.61}$$

where

$$r_0 \equiv 1/(\xi - q) \qquad \text{and} \qquad \lambda_0 \equiv \cos \eta. \tag{3.62}$$

Inserting (3.60) and (3.61) in (3.48)–(3.50) we easily establish that, to the same order of accuracy,

$$h_1(\xi, \eta) = r_0^2 \left(1 + q \sum_{j=2}^{4} (j + 2)\, r_0^{j+1} P_j(\lambda_0)\right), \tag{3.63}$$

$$h_2(\xi, \eta) = r_0 \left(1 + q \sum_{j=2}^{4} \frac{r_0^{j+1}}{j + 1} [(j + 1)(j + 2)\, P_j(\lambda_0) - P'_{j+1}(\lambda_0)]\right), \tag{3.64}$$

$$h_3(\xi, \eta) = r_0 \sqrt{1 - \lambda_0^2} \left(1 + q \sum_{j=2}^{4} \frac{r_0^{j+1}}{j + 1} P'_{j+1}(\lambda_0)\right); \tag{3.65}$$

and the direction cosines (3.51)–(3.59) can be rewritten in terms of ξ and η with the aid of (3.60) and (3.61) if needed with equal ease.

C. *Double-Star Problem*

In the preceding two sections of this part of our chapter we established the existence and approximate form of Roche curvilinear coordinates associated with the separate effects of rotation and tides. In actual double stars the effects of rotational and tidal distortion are, however, of the same order of magnitude, and their first-order effects are additive. Therefore, the Roche coordinates associated with the components of close binary systems revolving around the common center of gravity should be obtained by an integration of the complete system (3.1)–(3.3) of orthogonality conditions for the Roche potential ξ as given by Eq. (2.5).

In Sections III,A and III,B we integrated this system for the separate cases of $q = 0$, $n > 0$ and $n = 0$, $q > 0$. In each case, one Roche surface associated with the angular coordinates turned out to be a plane—a fact which reduced the order of the differential system (3.1)–(3.3) by one, and reduced the corresponding set of curvilinear coordinates to constitute a Lamé family. In the general double-star case this will, however, no longer be true: the surfaces over which all three orthogonal coordinates $\xi, \eta, \zeta =$ constant will be curved; and the question arises whether or not, under these conditions, the differential system (3.1)–(3.3) continues to be integrable for the general form of $\xi(r, \theta, \phi)$ as given by Eq. (2.5).

In order to answer the question, let us return to Eqs. (3.3)–(3.3) and eliminate between them the derivatives of either η or ζ (on account of the symmetry, the choice is immaterial). Eliminating η we obtain the following determinantal equation.

$$\begin{vmatrix} \xi_x & \zeta_x & \zeta_x\xi_{xx} + \zeta_y\xi_{xy} + \zeta_z\xi_{xz} \\ \xi_y & \zeta_y & \zeta_x\xi_{xy} + \zeta_y\xi_{yy} + \zeta_z\xi_{yz} \\ \xi_z & \zeta_z & \zeta_x\xi_{xz} + \zeta_y\xi_{yz} + \zeta_z\xi_{zz} \end{vmatrix} = 0 \tag{3.66}$$

between first and second partial derivatives of ξ and ζ, to which we should adjoin the relation (3.2) between first derivatives of the same variables.

The foregoing equation (3.66) is homogeneous and quadratic in $\zeta_{x,y,z}$. We can, therefore, solve (3.66) together with (3.2) for the *ratios* of $\zeta_{x,y,z}$ in the form

$$\zeta_x : \zeta_y : \zeta_z = L : M : N, \tag{3.67}$$

where L, M, N are functions, not only of the first derivatives of ξ as given by Eqs. (2.27)–(2.29), but also of its second derivatives (2.63)–(2.68).

In order that the function $\zeta(x, y, z)$ be integrable from (3.1) to (3.3), it is necessary (cf. Darboux, 1910; Forsyth, 1912) that the total differential equation

$$L\,dx + M\,dy + N\,dz = 0 \tag{3.68}$$

be integrable; and its condition of integrability is known to be of the form

$$L(M_z - N_y) + M(N_x - L_z) + N(L_y - M_x) = 0, \tag{3.69}$$

where the subscripts x, y, z denote again partial differentiation with respect to the respective variable.

If a family of surfaces represented by the equation $\xi(x, y, z) = \text{constant}$ be capable of generating an integrable triply orthogonal system, it is both necessary and sufficient that the generating function ξ should satisfy the above partial differential equation (3.69) of fourth degree and third order. An evaluation of the left-hand side of (3.69) for ξ as given by (2.5) was recently undertaken by Kopal and Ali (1971). The result disclosed that this equation is indeed satisfied if either n or q are zero (i.e., in the case of pure rotational or tidal distortion), but *not* if both n and q are nonvanishing—as they indeed are in the double-star problem. Therefore, the general Roche potential of the form (2.5)—which constitutes a scalar sum of the effects arising from rotation and tides—renders the orthogonality conditions (3.1)–(3.3) nonintegrable analytically because it fails to satisfy the integrability condition (3.69).

It may be noted that the situation we meet is analogous to that encountered in the restricted problem of three bodies, of which ξ represents indeed the potential. As is well known, for $q = 0$ and $n > 0$, the three-body problem reduces to that of two bodies which is analytically integrable in terms of elementary functions; while if $n = 0$ and $q > 0$, the three-body problem reduces
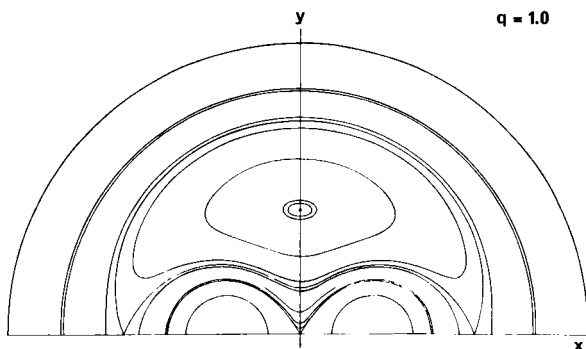
FIG. 3. Cross sections of a family of Roche equipotentials ξ = constant with the xy-plane for $y \geqslant 0$, corresponding to 12 different values of the potential and the mass-ratio $q = 1$ (after Kopal, 1959).
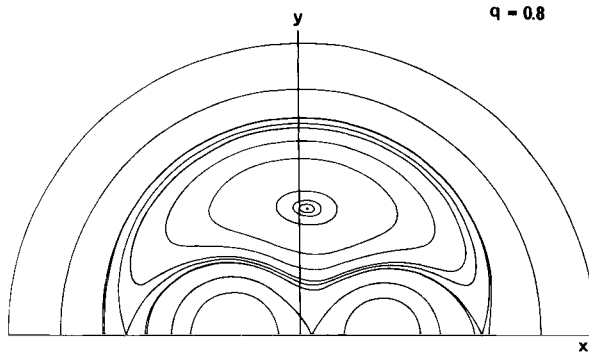
FIG. 4. Cross sections of a family of Roche equipotentials ξ = constant with the xy-plane for $y \geqslant 0$, corresponding to 11 different values of the potential, and the mass-ratio $q = 0.8$ (after Kopal, 1959).

to a "two-center" case which can likewise be integrated in terms of elliptic functions. However, since the days of Poincaré (1892) we know that if both n and q are different from zero, the restricted problem of three bodies does *not* admit of any integral (algebraic or transcendental) other than that of the energy—of which our Roche potential (2.5) represents a constant velocity surface.

It appears, therefore, that when we set out to construct a triply orthogonal system of curvilinear coordinates of which the complete Roche potential (2.5) of the double-star problem is a member, the fact that this form of ξ fails to fulfil Eq. (3.69) precludes the existence of any analytic integrals of our orthogonality conditions (3.1)–(3.3) other than ξ in the same way as Poincaré's theorem has ruled out the existence of any nontrivial integrals in the restricted
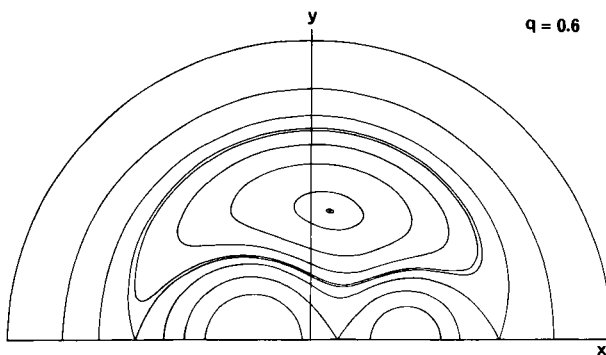
FIG. 5. Cross sections of a family of the Roche equipotentials ξ = constant with the xy-plane for $y \geqslant 0$, corresponding to 10 different values of the potential, and the mass-ratio $q = 0.6$ (after Kopal, 1959).
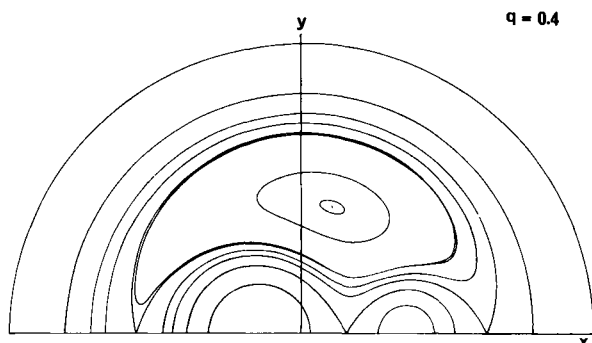
FIG. 6. Cross sections of a family of the Roche equipotentials ξ = constant with the xy-plane for $y \geqslant 0$, corresponding to 8 different values of the potential, and the mass-ratio $q = 0.4$ (after Kopal, 1959).

problem of three bodies other than the energy integral; and numerical integrations represent the only avenue of approach towards their construction.

Extensive and accurate numerical integrations of the system (3.1)–(3.3) have indeed been recently carried out by Kitamura (1970) in such cases when this system can be reduced to ordinary differential equations. Such a possibility
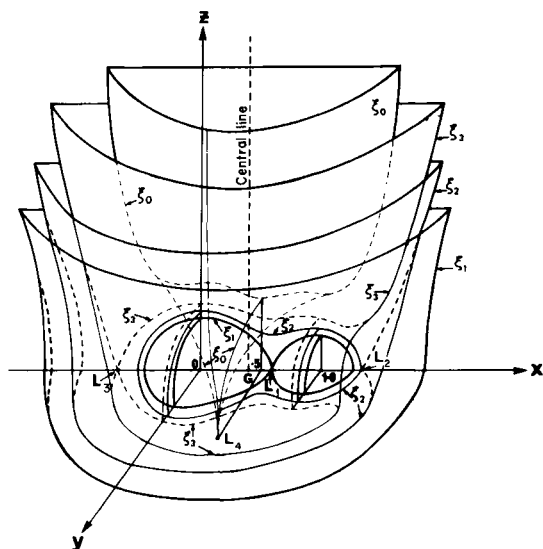
FIG. 7. A schematic view of the Roche equipotentials ξ = constant in three-dimensional perspective, corresponding to the values which the potential assumes at different Lagrangian points of the configuration (after Kitamura, 1970).
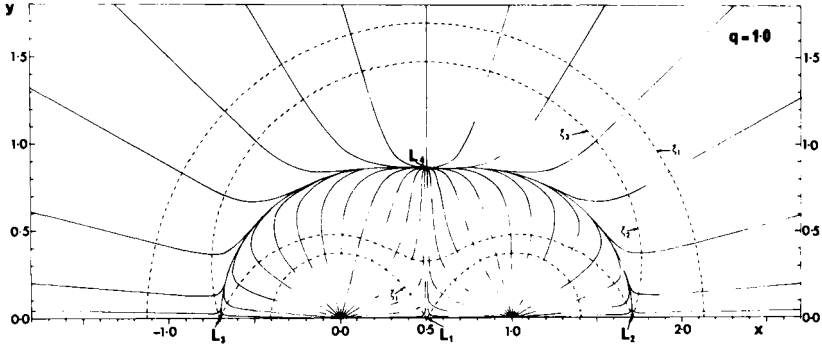
FIG. 8. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 1$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).

arises indeed in the planes $x = 0$, $y = 0$, or $z = 0$ where the families of the surfaces defined by ξ = constant, η = constant, or ζ = constant intersect the Cartesian planes in the form of two-dimensional curves, which can be generated by numerical solutions of ordinary differential equations in two variables.

Before we do so, however, let us consider in more detail the geometrical characteristics of the curves obtaining by an intersection of the surfaces $\xi(x, y, z)$ with the xy-plane. Figure 1 has given us a schematic sketch of such curves; and a more detailed presentation of the profiles of the functions $\xi(x, y, 0)$
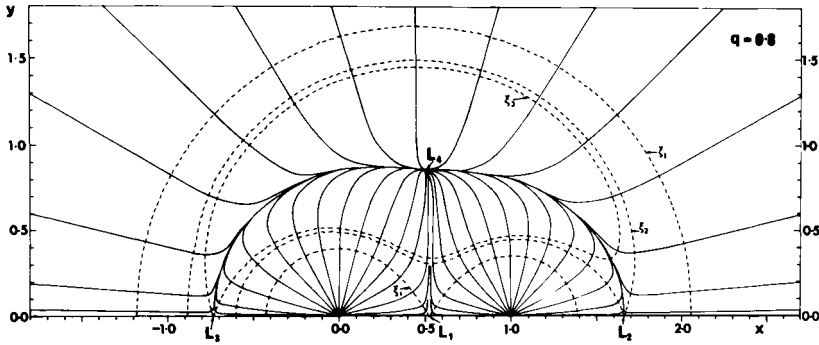
FIG. 9. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 0.8$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
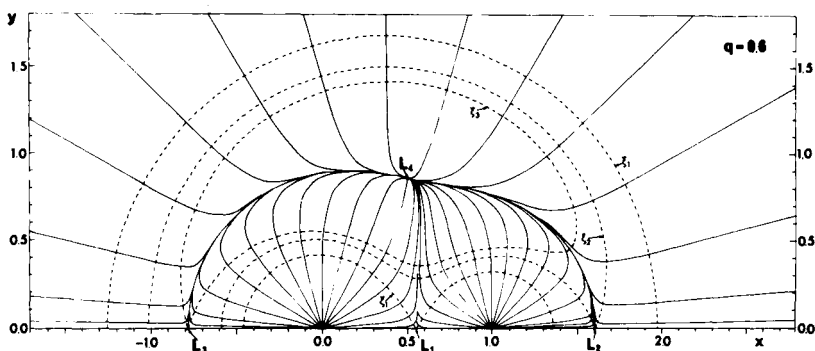
FIG. 10. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 0.6$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).

in the plane $z = 0$ can be found in the accompanying Figures 3 to 6, drawn to scale for four values of the mass-ratio $q = 1.0\,(-0.2)\,0.4$; while Fig. 7 contains a sketch of a three-dimensional representation of the surfaces $\xi(x, y, z) = \xi_{1,2,\ldots,5}$ passing through the five Lagrangian points discussed in Section II,E.

Moreover, in the same plane $(z = 0) = 0$; and if so, the orthogonality conditions (3.1)–(3.3) reduce to the single requirement that

$$\xi_x \eta_x + \xi_y \eta_y = 0, \tag{3.70}$$
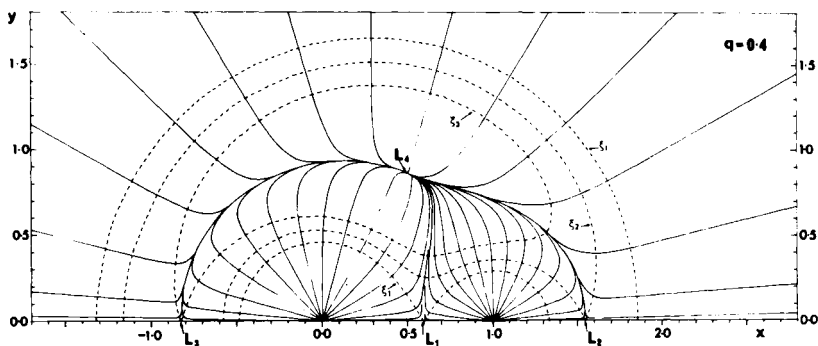
FIG. 11. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 0.4$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
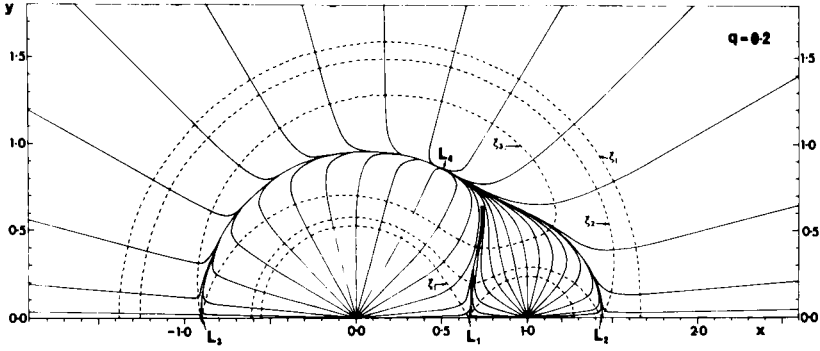
FIG. 12. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 0.2$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).

disclosing that the curves $\eta(x, y)$ = constant in the xy-plane should be generated by one-parameter solutions of the ordinary differential equation

$$\frac{dx}{dy} = \frac{\xi_x}{\xi_y} = \frac{x(1 - r_1^3)\, r_2^3 - q(1 - x)\, r_1^3(1 - r_2^3)}{y[(1 - r_1^3)\, r_2^3 + q r_1^3(1 - r_2^3)]}, \tag{3.71}$$

where

$$r_1^2 = x^2 + y^2 \quad \text{and} \quad r_2^2 = (1 - x)^2 + y^2. \tag{3.72}$$

In the immediate proximity of the origin, approximate analytical solutions of the linearized equation (3.70) were constructed by Kopal (1969); but it was
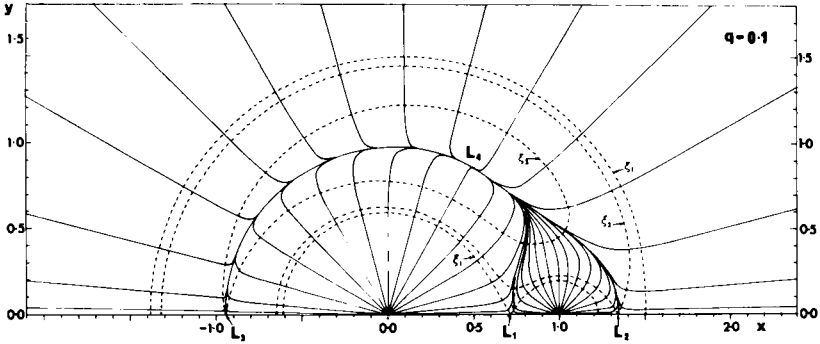
FIG. 13. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 0.1$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
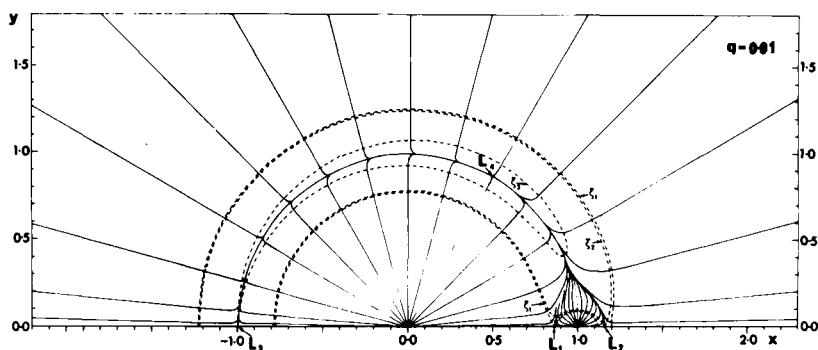
FIG. 14. The geometry of the curves η = constant in the xy-plane for $y \geqslant 0$, corresponding to the mass-ratio $q = 0.01$. The trajectories of $\eta(x, y, 0)$ are marked with full curves; while the dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
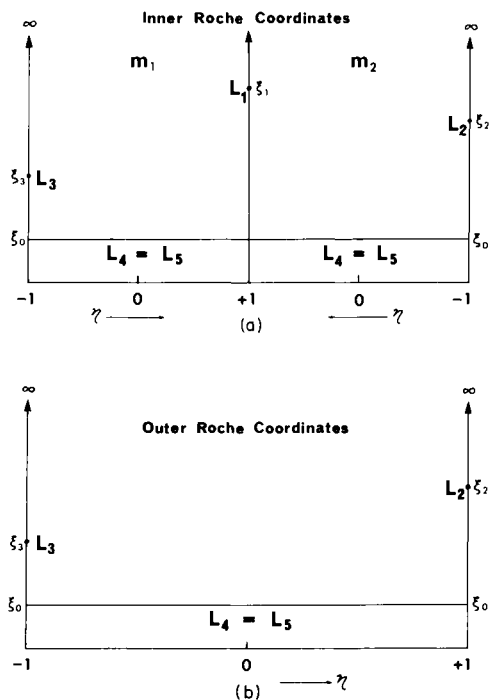
FIG. 15. Mapping of the positions of the Lagrangian points $L_{1,\ldots,5}$ of the xy-plane on to the $\xi\eta$-plane in the Roche curvilinear coordinates (a) within, and (b) outside, the envelope of the η = constant curves (after Kitamura, 1970).
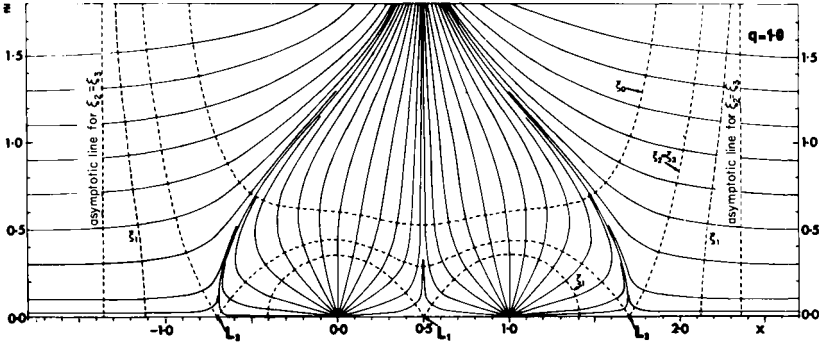
FIG. 16. The geometry of the curves ζ = constant in the xz-plane for $z \geqslant 0$, corresponding to the mass-ratio $q = 1$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).

not until Kitamura (1970) embarked on extensive numerical integration of the exact (nonlinear) equation (3.71) that the topological properties of the family of curves defined by the equation $(x, y, 0)$ = constant emerged from his work, and their principal features can be summarized as follows.

For large values of ξ—in close proximity of the two finite centers of mass 1 and q—the curves η = constant depart but slightly from straight lines passing through the two centers; and so they are at great distances from our gravitational dipole, where they are tangent to the quasicircular potential "curtains" surrounding the system from outside. In the intermediate regions of the plane
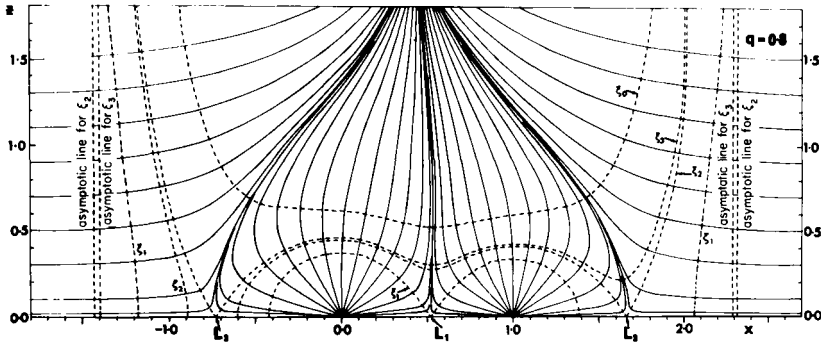
FIG. 17. The geometry of the curves ζ = constant in the xz-plane for $z \geqslant 0$, corresponding to the mass-ratio $q = 0.8$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
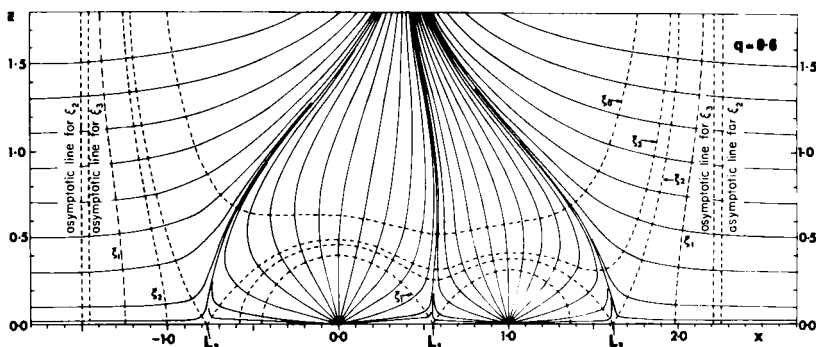
FIG. 18. The geometry of the curves ζ = constant in the xz-plane for $z \geqslant 0$, corresponding to the mass-ratio $q = 0.6$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).

$Z = 0$, however, both the inner and outer tangent lines curve along a common envelope, and pass through the Lagrangian equilateral triangle points $L_{4,5}$. However, the positions of the collinear Lagrangian points $L_{1,2,3}$ can be reached with the limiting η = constant lines which coincide with the x- or y-axis. This remains true for every mass-ratio $0 \leqslant q \leqslant 1$.

These results make it evident that, in the immediate proximity of the mass-centers of the two components—or at a sufficiently great distance from them—the $\xi\eta$-Roche coordinates tend asymptotically to plane-polar coordinates in which ξ plays the role of the radial (and η, of the angular) coordinate. However,
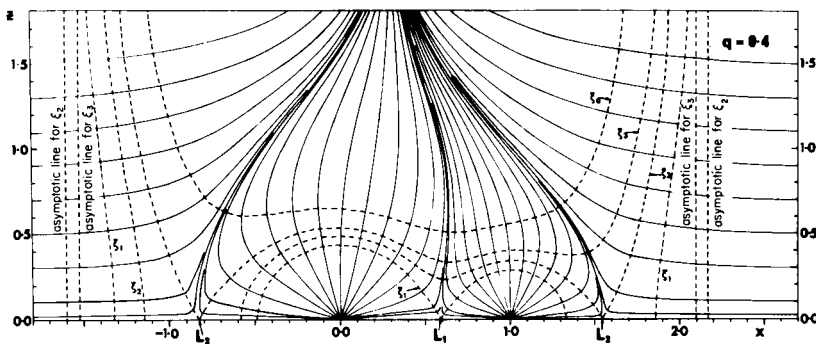
FIG. 19. The geometry of the curves ζ = constant in the xz-plane for $z \geqslant 0$, corresponding to the mass-ratio $q = 0.4$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
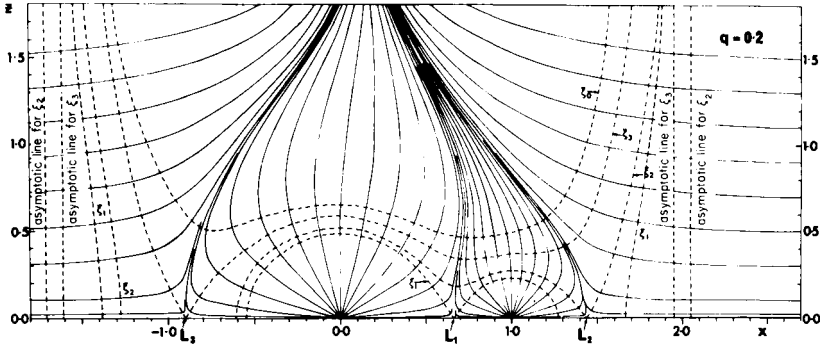
FIG. 20. The geometry of the curves ζ = constant in the xz-plane for $z \geqslant 0$, corresponding to the mass-ratio $q = 0.2$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).

in the intermediate region the situation becomes very much more complicated, as shown in Figs. 8–14 representing, in graphical form, the results of Kitamura's numerical integrations of Eq. (3.71) for seven different values of the mass-ratio q ranging from 1 to 0.01.

The mapping of the principal features in the xy-plane on to the $\xi\eta$-plane is illustrated in the accompanying figure 15. The mass-points m, m', as well as the potential "curtain" surrounding the whole system, will be represented on the $\xi\eta$-plane by the abscissae $\xi = \infty$. Closed ovals surrounding m, m' will be represented by the abscissae $\xi > \xi_1$, where ξ_1 stands for the (normalized) value of the potential at which the closed loops surrounding m and m' coalesce
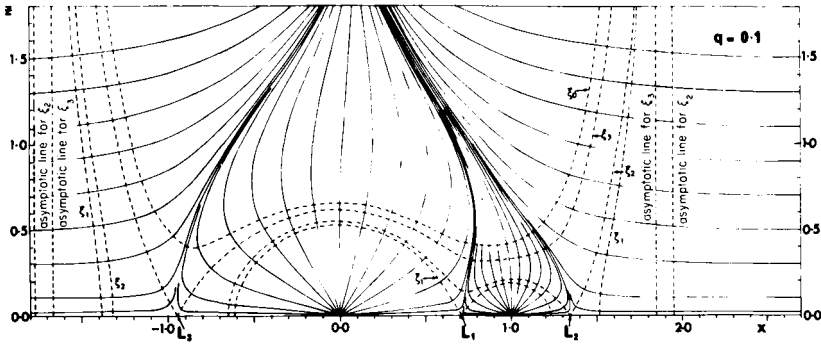
FIG. 21. The geometry of the curves ζ = constant in the xz-plane for $z \geqslant 0$, corresponding to the mass-ratio $q = 0.1$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
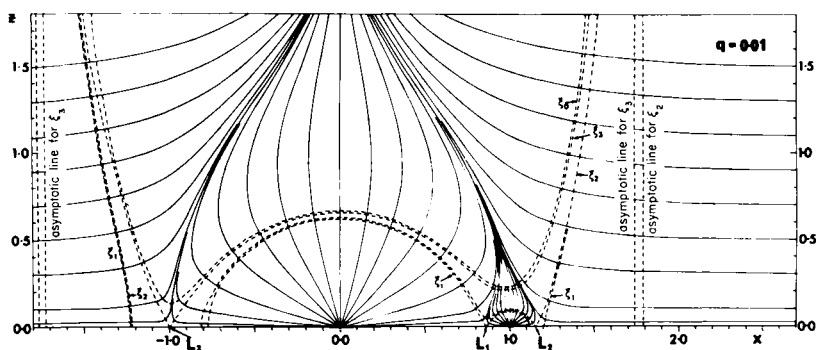
FIG. 22. The geometry of the curves ζ = constant in the xz-plane for $z \geq 0$, corresponding to the mass-ratio $q = 0.01$. The trajectories of $\zeta(x, 0, z)$ are marked with full curves; while dotted curves represent the equipotentials ξ = constant for the values which the potential assumes at the Lagrangian collinear points $L_{1,2,3}$ (after Kitamura, 1970).
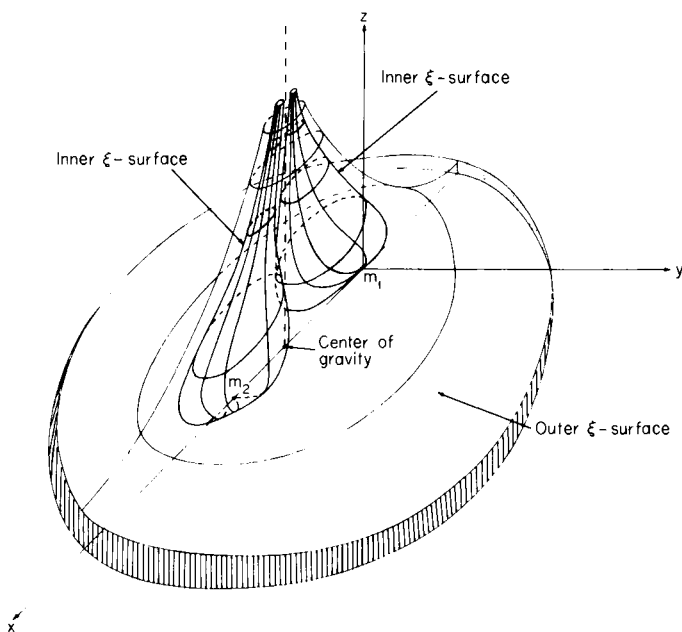
FIG. 23. Three-dimensional representation of the surfaces $\zeta(x, y, z)$ = constant (after Kitamura, 1970).

at the inner Lagrangian point L_1, the position of which is specified on the upper part of Fig. 15 by $\xi = \xi_1$ and $\eta = 0°$. For $\xi < \xi_1$ (lower part of the diagram) the abscissae $\xi =$ constant will represent dumbbell figures seen in the xy-plane on Figs. 3 to 6, which eventually open up at the Lagrangian points $L_2(\xi_2, 0°)$ and $L_3(\xi_3, \pi)$; while the abscissa ξ_0 represents the location on our diagram of the equilateral points $L_{4,5}$.

If we turn now to the plane $y = 0$, the curves $(x, 0, z)$ in the xz-plane must satisfy the orthogonality condition

$$\xi_x \zeta_x + \xi_z \zeta_z = 0, \tag{3.73}$$

and as such will be generated by the solutions of the first-order nonlinear differential equation

$$\frac{dx}{dz} = \frac{\xi_x}{\xi_z} = \frac{x(1 - r_1^3) r_2^3 - q(1 - x) r_1^3(1 - r_2^3)}{z(1 + qr_1^3)}, \tag{3.74}$$

with

$$r_1^2 = x^2 + z^2 \quad \text{and} \quad r_2^2 = (1 - x)^2 + z^2. \tag{3.75}$$

This equation has likewise been integrated extensively by Kitamura (1970) numerically, and the profiles of the corresponding curves $\zeta(x, 0, z)$ in the xz-plane are shown in the accompanying Figs. 16–22; while Fig. 23 gives a representation of the surfaces $\zeta(x, y, z)$ in three-dimensional perspective based on numerical integrations.

IV

STABILITY OF THE ROCHE MODEL

In the preceding part of this chapter we have developed an orthogonal system of curvilinear coordinates based on the Roche equipotentials, and pointed out the advantages for problems encountered in connection with close binary systems—in particular, the simple definition of the distorted surface by the equation $\xi =$ constant, for which the curvature of the surfaces $\eta =$ constant or $\zeta =$ constant is a relatively small price to pay.

The aim of the present part of this chapter will be to illustrate further the use of the Roche coordinates by an application to the problems of the *stability* of distorted components of close binary systems, or to problems of fluid flow of gas streams in close binaries for which the Roche equipotentials constitute boundary conditions. With the metric coefficients $h_{1,2,3}$ specifying the transformation of an arbitrary line element (3.4) from the Cartesian to Roche coordinates established in the preceding Section III of this chapter, we are now in a position to rewrite the fundamental equations of hydrodynamics in

terms of the curvilinear coordinates ξ, η, ζ in the form in which they can be used for such work.

In what follows, let $u_{\xi,\eta,\zeta}$ denote the curvilinear components of the velocity vector in the directions of increasing ξ, η, ζ by the equations

$$u_\xi = h_1\dot{\xi} = l_1\dot{x} + m_1\dot{y} + n_1\dot{z}, \tag{4.1}$$

$$u_\eta = h_2\dot{\eta} = l_2\dot{x} + m_2\dot{y} + n_2\dot{z}, \tag{4.2}$$

$$u_\zeta = h_3\dot{\zeta} = l_3\dot{x} + m_3\dot{y} + n_3\dot{z}, \tag{4.3}$$

where dots stand for the time derivatives of the respective variables, and the direction cosines l_j, m_j, n_j ($j = 1, 2, 3$) continue to be defined by Eqs. (3.8)–(3.10).

If so, the fundamental equation of hydrodynamics in the Eulerian form can be rewritten in the scalar form as

$$h_1\frac{Du_\xi}{Dt} + \frac{1}{\rho}\frac{\partial P}{\partial \xi} = \frac{\partial \Omega}{\partial \xi}, \tag{4.4}$$

$$h_2\frac{Du_\eta}{Dt} + \frac{1}{\rho}\frac{\partial P}{\partial \eta} = 0, \tag{4.5}$$

$$h_3\frac{Du_\zeta}{Dt} + \frac{1}{\rho}\frac{\partial P}{\partial \zeta} = 0, \tag{4.6}$$

where t denotes the time, P the pressure inside the fluid, ρ the density, and Ω the total potential of forces acting upon any element of the fluid, and where (see, e.g., Lamb, 1932) the Lagrangian components of the accelerations on the left-hand sides of Eqs. (4.4)–(4.6) in the Roche curvilinear coordinates are expressible as

$$\begin{aligned}\frac{Du_\xi}{Dt} = {} & \frac{\partial u_\xi}{\partial t} + \frac{u_\xi}{h_1}\frac{\partial u_\xi}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial u_\xi}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial u_\xi}{\partial \zeta} + \frac{u_\xi}{h_1}\left(\frac{u_\xi}{h_1}\frac{\partial h_1}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial h_1}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial h_1}{\partial \zeta}\right) \\ & - \frac{1}{h_1}\left(\frac{u_\xi^2}{h_1}\frac{\partial h_1}{\partial \xi} + \frac{u_\eta^2}{h_2}\frac{\partial h_2}{\partial \xi} + \frac{u_\zeta^2}{h_3}\frac{\partial h_3}{\partial \xi}\right),\end{aligned} \tag{4.7}$$

$$\begin{aligned}\frac{Du_\eta}{Dt} = {} & \frac{\partial u_\eta}{\partial t} + \frac{u_\xi}{h_1}\frac{\partial u_\eta}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial u_\eta}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial u_\eta}{\partial \zeta} + \frac{u_\eta}{h_2}\left(\frac{u_\xi}{h_1}\frac{\partial h_2}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial h_2}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial h_2}{\partial \zeta}\right) \\ & - \frac{1}{h_2}\left(\frac{u_\xi^2}{h_1}\frac{\partial h_1}{\partial \eta} + \frac{u_\eta^2}{h_2}\frac{\partial h_2}{\partial \eta} + \frac{u_\zeta^2}{h_3}\frac{\partial h_3}{\partial \eta}\right),\end{aligned} \tag{4.8}$$

$$\begin{aligned}\frac{Du_\zeta}{Dt} = {} & \frac{\partial u_\zeta}{\partial t} + \frac{\mu_\xi}{h_1}\frac{\partial u_\zeta}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial u_\zeta}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial u_\zeta}{\partial \zeta} + \frac{u_\zeta}{h_3}\left(\frac{u_\xi}{h_1}\frac{\partial h_3}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial h_3}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial h_3}{\partial \zeta}\right) \\ & - \frac{1}{h_3}\left(\frac{u_\xi^2}{h_1}\frac{\partial h_1}{\partial \zeta} + \frac{u_\eta^2}{h_2}\frac{\partial h_2}{\partial \zeta} + \frac{u_\zeta^2}{h_3}\frac{\partial h_3}{\partial \zeta}\right).\end{aligned} \tag{4.9}$$

As is well known, the Eulerian equations (4.4)–(4.6) of motion safeguard the conservation of ***momentum*** of the respective dynamical system. To safeguard the conservation of ***mass***, we must invoke the use of the equation of continuity

$$\frac{D\rho}{Dt} + \rho\Delta = 0, \tag{4.10}$$

where

$$\Delta = \frac{1}{h_1 h_2 h_3}\left(\frac{\partial}{\partial \xi}(h_2 h_3 u_\xi) + \frac{\partial}{\partial \eta}(h_3 h_1 u_\eta) + \frac{\partial}{\partial \zeta}(h_1 h_2 u_\zeta)\right) \tag{4.11}$$

stands for the divergence of the velocity vector, and which represents a single scalar equation between the variables already involved in (4.4)–(4.6).

Equations (4.4)–(4.6) and (4.10) constitute, in fact, a simultaneous system of four differential equations for five dependent variables $u_{\xi,\eta,\zeta}$, P, and ρ which is as yet not sufficient for a complete specification of any one of them. In order to render our problem determinate, we must adjoin to it one additional equation relating the same variables; and this can be done by expressing the conservation of ***energy*** in a form appropriate for the problem. If, in particular, this energy can be conserved with the aid of an "equation of state" requiring the changes of P and ρ to be adiabatic, the fifth equation needed to complete our system is known to assume the form

$$DP/Dt = a^2\, D\rho/Dt, \tag{4.12}$$

where

$$a^2 = \gamma\, P/\rho \tag{4.13}$$

denotes the square of the local velocity of sound (γ being the ratio of specific heats at constant pressure and volume of the respective fluid). Moreover, inasmuch as the pressure P and density ρ are scalar quantities, their Lagrangian time derivatives can be expressed by

$$\frac{D}{Dt}(P, \rho) = \left(\frac{\partial}{\partial t} + \frac{u_\xi}{h_1}\frac{\partial}{\partial \xi} + \frac{u_\eta}{h_2}\frac{\partial}{\partial \eta} + \frac{u_\zeta}{h_3}\frac{\partial}{\partial \zeta}\right)(P, \rho); \tag{4.14}$$

and on insertion of (4.14) in (4.10) and (4.12) a specification of the equations of our problem in terms of the Roche coordinates ξ, η, ζ is now complete.

A. *Vibrations of the Roche Model*

The general equations of hydrodynamics in Roche coordinates offer considerable advantages over their Cartesian form for treatment of many problems—in particular, of those in which the boundary conditions can be simplified in their terms. For example, any specific property (such as the constancy of pressure

or density, or the vanishing of the gradients of any one of the dependent variables) to be enforced over the distorted surface of a star can be localized with the aid of only one of the three Roche spatial coordinates (e.g., ξ). In particular, to describe the free surface of a distorted component in a close binary system in terms of rectangular or polar coordinates may be a matter of some complexity (cf. Eq. 2.18); but in the Roche coordinates it is sufficient to do so by setting $\xi =$ constant (say, ξ_0). Or, in problems connected with fluid flow of gas in such systems the boundary conditions requiring the vanishing of motion normal to the surface of a star reduce to $u_\xi(\xi_0) = 0$.

The aim of the present section, however, will be to demonstrate such advantages on one particular example: namely, on the problem of *vibrational stability* of self-gravitating gaseous configurations, the surface of which can be represented by Roche equipotentials (2.5). Suppose that a configuration of this shape becomes disturbed—by whatever process—from its state of equilibrium; and its consequent change in size or form will render its dependent variables functions of the time. Under which conditions can such motions remain periodic?

In order to embark on an effort to answer this question, assume first that all effects caused by the initial disturbance are small enough for quantities of the order of their squares and higher powers to be ignorable. Second, let us assume that the pressure P and density ρ can be expressed as

$$P(\xi, \eta, \zeta; t) = P_0(\xi) + P'(\xi, \eta, \zeta; t), \tag{4.15}$$

$$\rho(\xi, \eta, \zeta; t) = \rho_0(\xi) + \rho'(\xi, \eta, \zeta; t), \tag{4.16}$$

where P_0 and ρ_0 describe the respective properties of our configuration in its equilibrium state (depending on ξ only); while P', ρ' represent their changes brought about by motion with the velocity components $u_{\xi,\eta,\zeta}$. If, moreover, P', ρ' as well as $u_{\xi,\eta,\zeta}$ can be regarded as small quantities of first order, our fundamental system of Eulerian equations (4.4)–(4.6) can be *linearized* to yield

$$h_1 \frac{\partial u_\xi}{\partial t} + \frac{1}{\rho_0}\frac{\partial P_0}{\partial \xi} - \frac{\partial \Omega}{\partial \xi} + \frac{1}{\rho_0}\left(\frac{\partial P'}{\partial \xi} - \frac{\rho'}{\rho_0}\frac{\partial P_0}{\partial \xi}\right) = 0, \tag{4.17}$$

$$h_2 \frac{\partial u_\eta}{\partial t} + \frac{1}{\rho_0}\frac{\partial P_0}{\partial \eta} + \frac{1}{\rho_0}\left(\frac{\partial P'}{\partial \eta} - \frac{\rho'}{\rho_0}\frac{\partial P_0}{\partial \eta}\right) = 0, \tag{4.18}$$

$$h_3 \frac{\partial u_\zeta}{\partial t} + \frac{1}{\rho_0}\frac{\partial P_0}{\partial \zeta} + \frac{1}{\rho_0}\left(\frac{\partial P'}{\partial \zeta} - \frac{\rho'}{\rho_0}\frac{\partial P_0}{\partial \zeta}\right) = 0. \tag{4.19}$$

Since, moreover, consistent with (4.15) and (4.16)

$$\frac{\partial P_0}{\partial \eta} = \frac{\partial P_0}{\partial \zeta} = 0 \tag{4.20}$$

and, in the equilibrium state,

$$\frac{\partial P_0}{\partial \xi} = \rho_0 \frac{\partial \Omega}{\partial \xi} = k\rho_0 , \tag{4.21}$$

where the constant k can be normalized to 1, the foregoing system (4.17)–(4.19) of linearized Eulerian equations in Roche coordinates readily reduces to

$$\rho_0 h_1 \frac{\partial u_\xi}{\partial t} = \rho' - \frac{\partial P'}{\partial \xi} , \tag{4.22}$$

$$\rho_0 h_2 \frac{\partial u_\eta}{\partial t} = - \frac{\partial P'}{\partial \eta} , \tag{4.23}$$

$$\rho_0 h_3 \frac{\partial u_\zeta}{\partial t} = - \frac{\partial P'}{\partial \zeta} ; \tag{4.24}$$

while the linearized equations of continuity (4.10) and of adiabatic changes of state (4.12) similarly reduce to

$$\frac{\partial P'}{\partial t} + \frac{u_\xi}{h_1} \frac{\partial \rho_0}{\partial \xi} + \rho_0 \Delta = 0 \tag{4.25}$$

and

$$\frac{\partial P'}{\partial t} + \frac{u_\xi}{h_1} \rho_0 + a^2 \rho_0 \Delta = 0, \tag{4.26}$$

respectively.

In order to eliminate P' and ρ' from Eqs. (4.22)–(4.24), differentiate the latter with respect to t and insert for $\partial P'/\partial t$ and $\partial \rho'/\partial t$ from (4.25) and (4.26); the result can be expressed as

$$\frac{\partial^2}{\partial t^2}(h_1 u_\xi) = \frac{\partial}{\partial \xi}\left(a^2 \Delta + \frac{u_\xi}{h_1}\right) + a^2 A \Delta, \tag{4.27}$$

$$\frac{\partial^2}{\partial t^2}(h_2 u_\eta) = \frac{\partial}{\partial \eta}\left(a^2 \Delta + \frac{u_\xi}{h_1}\right), \tag{4.28}$$

$$\frac{\partial^2}{\partial t^2}(h_3 u_\zeta) = \frac{\partial}{\partial \zeta}\left(a^2 \Delta + \frac{u_\xi}{h_1}\right), \tag{4.29}$$

where (consistent with 4.13) $a^2 = \gamma P_0/\rho_0$, and where we have abbreviated

$$A = \frac{1}{\rho_0} \frac{\partial \rho_0}{\partial \xi} - \frac{1}{\gamma P_0} \frac{\partial P_0}{\partial \xi} = \frac{\partial}{\partial \xi} \log \rho_0 P_0^{-1/\gamma}; \tag{4.30}$$

in a linearized problem both a^2 and A are functions of ξ alone.

If, moreover, the motion we anticipate is to be *harmonic*, such that

$$\partial^2/\partial t^2 \equiv -\tilde{\nu}^2, \tag{4.31}$$

where $\tilde{\nu}^2$ denotes a positive frequency of oscillatory motion, our system (4.27)–(4.29) further simplifies to

$$\Psi_\xi + \tilde{\nu}^2 h_1 u_\xi + a^2 A\Delta = 0, \tag{4.32}$$

$$\Psi_\eta + \tilde{\nu}^2 h_2 u_\eta = 0, \tag{4.33}$$

$$\Psi_\zeta + \tilde{\nu}^2 h_3 u_\zeta = 0, \tag{4.34}$$

where we have abbreviated

$$\Psi \equiv a^2\Delta + k_1^{-1} u_\xi , \tag{4.35}$$

and the divergence Δ continues to be given by Equation (4.11).

Accordingly, the foregoing system (4.32)–(4.34) consists of three simultaneous partial differential equations in ξ, η, ζ of second order; so that the problem represented by them is one of *sixth* order. However, if the distortion of the Roche equipotentials ξ is caused by either rotation or tides alone, it is easy to show that this order becomes diminished to *four*. Thus in the case of distortion caused by axial rotation with constant angular velocity $\omega^2 = n$, by (4.2)

$$u_\eta = \omega h_2 \tag{4.36}$$

and, therefore,

$$\partial u_\eta/\partial t = 0. \tag{4.37}$$

Since, moreover, P' does not depend on η in the case of rotational-spheroid symmetry, Eq. (4.33) is satisfied by both its sides vanishing identically—in which case the system (4.32)–(4.34) reduces to

$$\Psi_\xi + \tilde{\nu}^2 h_1 u_\xi + a^2 A\Delta = 0, \tag{4.38}$$

$$\Psi_\zeta + \tilde{\nu}^2 h_3 u_\zeta = 0, \tag{4.39}$$

where Ψ has already been defined by (4.35),

$$\Delta = \frac{\partial}{\partial\xi}\left(\frac{u_\xi}{h_1}\right) + \left(\frac{\partial}{\partial\xi}\log|\,h_1h_2h_3\,|\right)\frac{u_\xi}{h_1} + \frac{\partial}{\partial\zeta}\left(\frac{u_\zeta}{h_3}\right) + \left(\frac{\partial}{\partial\zeta}\log|\,h_1h_2h_3\,|\right)\frac{u_\zeta}{h_3}, \tag{4.40}$$

and where the metric coefficients $h_{1,2,3}(\xi, \zeta)$ in the neighborhood of the origin continue to be given by the expansions on the right-hand sides of Eqs. (3.38)–(3.40).

On the other hand, in the case of purely tidal distortion of a nonrotating configuration, it is the third equation (4.34) which vanishes identically, leaving us with the fourth-order system

$$\Psi_\xi + \tilde{\nu}^2 h_1 u_\xi + a^2 A \Delta = 0, \tag{4.41}$$

$$\Psi_\eta + \tilde{\nu}^2 h_2 u_\eta = 0, \tag{4.42}$$

where

$$\Delta \equiv \frac{\partial}{\partial \xi}\left(\frac{u_\xi}{h_1}\right) + \left(\frac{\partial}{\partial \xi}\log \mid h_1 h_2 h_3 \mid\right)\frac{u_\xi}{h_1} + \frac{\partial}{\partial \eta}\left(\frac{u_\eta}{h_2}\right) + \left(\frac{\partial}{\partial \eta}\log \mid h_1 h_2 h_3 \mid\right)\frac{u_\eta}{h_2}, \tag{4.43}$$

and where the metric coefficients $h_{1,2,3}(\xi, \eta)$ in the neighborhood of the origin are given by Eqs. (3.63)–(3.65).

For additional formal theorems concerning the stability of the Roche model as introduced in Section II,A of this chapter, cf. also Lichtenstein (1928) or Maruhn (1931).

B. *Roche Harmonics*

The differential equations established in the preceding section permit us to study the vibrational stability of centrally condensed stellar configurations distorted by rotation or tides to an acoustic approximation by ascertaining whether or not their particular solutions satisfying appropriate boundary conditions can be constructed—numerically or otherwise—for nonnegative values of $\tilde{\nu}^2$. The boundary conditions appropriate for our problem remain yet to be specified; and this we shall do in the next section. At this stage we wish, however, to introduce to the reader a family of new auxiliary functions, which we propose to call the *Roche harmonics*, generated by the solutions of the Laplace equation

$$\Delta^2 \Phi = 0 \tag{4.44}$$

in Roche curvilinear coordinates, and associated with the Roche equipotential surfaces (2.5) in the same way as the spherical harmonics are with a sphere. In point of fact, spherical harmonics should be regarded as a limiting case of more general Roche harmonics, for as $\xi \to \infty$, the latter must obviously reduce to the former in the immediate neighborhood of the origin.

An investigation of the formal mathematical properties of the Roche harmonics (such as their orthogonality, etc.) has not yet been undertaken, and lies wholly outside the scope of the present review article. However, in what follows we wish to give the explicit form of the first few such harmonics associated with

the rotational and tidal distortion of the Roche model, and to point out some of their more general properties on such examples.

In order to do so, let us depart from the well-known definition of the Laplacian operator

$$\nabla^2 = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial \xi}\left(\frac{h_2 h_3}{h_1}\frac{\partial}{\partial \xi}\right) + \frac{\partial}{\partial \eta}\left(\frac{h_1 h_3}{h_2}\frac{\partial}{\partial \eta}\right) + \frac{\partial}{\partial \zeta}\left(\frac{h_1 h_2}{h_3}\frac{\partial}{\partial \zeta}\right)\right] \qquad (4.45)$$

in curvilinear coordinates ξ, η, ζ, where $h_{1,2,3}$ represent the metric coefficients of the transformation (3.4). For Roche equipotentials ξ as given by Eq. (2.5) the functions $h_{1,2,3}(\xi, \eta, \zeta)$ defined by (3.5)–(3.7) are not known to us in a closed form; but analytical expressions approximating them in the case of rotational or tidal distortion have been obtained by us in Sections III,A and B.

Making use of Eqs. (3.38)–(3.40), valid for the case of purely *rotational distortion,* we find the explicit form of the Laplace equation (4.44) becomes

$$\begin{aligned} &\left(1 - 8nr_0^3(1-\nu_0^2) + \frac{8}{3}n^2r_0^6(1-\nu_0^2)(9+13\nu_0^2) + \cdots\right)\frac{\partial}{\partial r_0}\left(r_0^2\frac{\partial \Phi}{\partial r_0}\right) - 4nr_0^4\frac{\partial \Phi}{\partial r_0} \\ &\quad + \left(1 - \frac{2}{3}nr_0^3(5-7\nu_0^2) - \frac{1}{9}n^2r_0^6(43 - 194\nu_0^2 + 143\nu_0^4) + \cdots\right)(1-\nu_0^2)\frac{\partial^2 \Phi}{\partial \nu_0^2} \\ &\quad - \left(1 + \frac{4}{3}nr_0^3\nu_0^2 + \frac{1}{9}n^2r_0^6(75 - 34\nu_0^2 - 33\nu_0^4) + \cdots\right)\left(2\nu_0\frac{\partial \Phi}{\partial \nu_0}\right) = 0, \end{aligned} \qquad (4.46)$$

correctly to terms of the order of $n^2 \equiv \frac{1}{4}\omega^4$, where

$$r_0 \equiv \xi^{-1} \quad \text{and} \quad \nu_0 \equiv \cos \zeta. \qquad (4.47)$$

If we seek its particular solution in ascending powers of r_0,[4] in the form of a series

$$\Phi = \sum_j a_j r_0^j R_j, \qquad (4.48)$$

where a_j are coefficients which, on account of the homogeneous nature of (4.46), remain arbitrary, and expand

$$R_j = P_j(\nu_0) + nr_0^3 X_2^{(j)}(\nu_0) + n^2 r_0^6 X_4^{(j)}(\nu_0) + \cdots, \qquad (4.49)$$

with $P_j(\nu_0)$ representing the well-known Legendre polynomials of ν_0, it can be

[4] The reader may note that should we seek solutions in descending powers of r_0, the equation $\nabla^2\Phi = n$ must be satisfied by $\Phi \equiv \xi = r_0^{-1}$ exactly, or to any approximation.

shown (cf. Kopal, 1971) that, to quantities of second order (i.e., of the order of r_0^8 on the right-hand side of 4.48), for $j > 1$,

$$X_2^{(2)}(\nu_0) = -(1 - \nu_0^2)(1 - 5\nu_0^2), \tag{4.50}$$

$$X_2^{(3)}(\nu_0) = -\tfrac{1}{2}(1 - \nu_0^2)(11 - 25\nu_0^2)\,\nu_0\,, \tag{4.51}$$

$$X_2^{(4)}(\nu_0) = \tfrac{1}{6}(1 - \nu_0^2)(9 - 120\nu_0^2 + 175\nu_0^4), \tag{4.52}$$

$$X_2^{(5)}(\nu_0) = \tfrac{5}{8}(1 - \nu_0^2)(17 - 98\nu_0^2 + 105\nu_0^4)\,\nu_0\,; \tag{4.53}$$

$$X_4^{(2)}(\nu_0) = -\tfrac{1}{6}(1 - \nu_0^2)(21 - 172\nu_0^2 + 175\nu_0^4), \tag{4.54}$$

etc.

The functions $R_j(\xi, \zeta)$ on the right-hand side of Eq. (4.49) represent the "Roche harmonics" which are associated with the rotational problem based on the potential function (3.11), and Eqs. (4.50)–(4.54) give their explicit form complete to terms of second order. In the case of purely *tidal distortion*, the metric coefficients $h_{1,2,3}(\xi, \eta)$ are known from Eqs. (3.63)–(3.65) only to quantities of first order in superficial distortion (i.e., of the order of qr_0^5). Making use of these, however, the Laplace equation (4.44) will assume the explicit form

$$\begin{aligned}\Big(1 - 2q \sum_{j=2}^{4} (j + 2)\, r_0^{j+1} P_j(\lambda_0)\Big)\Big[\frac{\partial}{\partial r_0}\Big(r_0^2 \frac{\partial \Phi}{\partial r_0}\Big) + \frac{\partial}{\partial \lambda_0}\Big((1 - \lambda_0^2)\frac{\partial \Phi}{\partial \lambda_0}\Big)\Big] \\ + 2q \sum_{j=2}^{4} \frac{r_0^{j+1}}{j+1}\Big[P'_{j+1}(\lambda_0)\frac{\partial}{\partial \lambda_0}\Big((1 - \lambda_0^2)\frac{\partial \Phi}{\partial \lambda_0}\Big) \\ + [(j + 2)\, P_j'(\lambda_0) - (j + 1)\, \lambda_0 P'_{j+1}(\lambda_0)]\frac{\partial \Phi}{\partial \lambda_0}\Big) = 0,\end{aligned} \tag{4.55}$$

where r_0 and λ_0 continue to be given by Eqs. (3.62).

If, moreover, we seek again a solution of (4.55) as a series of the form (4.48) in ascending powers of r_0,[5] and set

$$R_j = P_j(\lambda_0) + q \sum_{i=2}^{4} r_0^{i+1} Y_i^{(j)}(\lambda_0), \tag{4.56}$$

then (consistent with our adopted scheme of approximation) the functions $Y_i^{(j)}(\lambda_0)$ prove to be the polynomial solutions of the differential equation

$$\begin{aligned}\frac{\partial}{\partial \lambda_0}\Big((1 - \lambda_0^2)\frac{\partial Y_i^{(j)}}{\partial \lambda_0}\Big) + (i + j + 1)(i + j + 2)\, Y_i^{(j)} \\ + \frac{2}{i+1}((1 - \lambda_0^2)\, P''_{i+1}(\lambda_0)\, P''_j(\lambda_0) - j(j + 1)\, P_j(\lambda_0)\, P'_{j+1}(\lambda_0)) = 0,\end{aligned} \tag{4.57}$$

[5] In descending powers it is again satisfied exactly by $\Phi \equiv \xi$.

which are found to be of the form

$$Y_2^{(2)}(\lambda_0) = \tfrac{1}{2}(1 - 12\lambda_0^2 + 15\lambda_0^4), \tag{4.58}$$

$$Y_3^{(2)}(\lambda_0) = \tfrac{1}{8}(21 - 110\lambda_0^2 + 105\lambda_0^4)\,\lambda_0\,, \tag{4.59}$$

$$Y_4^{(2)}(\lambda_0) = -\,\tfrac{1}{8}(3 - 75\lambda_0^2 + 245\lambda_0^4 - 189\lambda_0^6); \tag{4.60}$$

$$Y_2^{(3)}(\lambda_0) = \tfrac{3}{4}(5 - 26\lambda_0^2 + 25\lambda_0^4)\,\lambda_0\,, \tag{4.61}$$

$$Y_3^{(3)}(\lambda_0) = -\,\tfrac{3}{16}(3 - 69\lambda_0^2 + 225\lambda_0^4 - 175\lambda_0^6); \tag{4.62}$$

$$Y_2^{(4)}(\lambda_0) = -\,\tfrac{1}{4}(3 - 69\lambda_0^2 + 225\lambda_0^4 - 175\lambda_0^6) = \tfrac{1}{4}Y_3^{(3)}(\lambda_0); \tag{4.63}$$

etc.

Equation (4.56) with (4.58)–(4.63) constitutes the explicit form of the "Roche harmonics" associated with first-order distortion of Roche tidal equipotential (3.41); but their generalization to quantities of second or higher orders still represents a task for the future.

C. *Stability near the Origin*

In the previous section of this chapter we derived the differential equations governing small oscillations of the Roche model in three dimensions. Although the linearized equations (4.38), (4.39) or (4.41), (4.42) of our problem can control such oscillations about the Roche figures of arbitrary distortion, we have not succeeded so far in constructing the explicit forms of the metric coefficients $h_{1.2.3}$ analytically in Sections III,A and B beyond the terms of first order in superficial distortion for the tidal problem, and those of second order for the rotational problem.

In the immediate proximity of the origin, however, our Roche equipotentials (2.5) tend asymptotically to become spherical (cf. Section II,A) for the rotational as well as tidal problem. In such a case, Eqs. (4.33) and (4.34) governing small oscillations of the Roche model will tend to be met by all their terms vanishing identically, while Eq. (4.32) reduces—on insertion of the first term from (4.40) or (4.43) for Δ—to the second-order differential equation

$$a^2 \frac{\partial^2 u_\xi}{\partial \xi^2} + \left(1 + a^2 A + \frac{\partial a^2}{\partial \xi}\right) \frac{\partial u_\xi}{\partial \xi} + \left[a^2 \frac{\partial H}{\partial \xi} + \left(\frac{\partial a^2}{\partial \xi} + a^2(A - H) - 1\right) H + \tilde{\nu}^2 h_1^2\right] u_\xi = 0, \tag{4.64}$$

where, in the immediate neighborhood of the origin,

$$h_1 = -r_0^2 + \cdots \qquad \text{and} \qquad H \equiv \frac{1}{h_1}\frac{\partial h_1}{\partial \xi} = -2r_0 + \cdots, \tag{4.65}$$

r_0 denoting (in accordance with Eqs. 3.36 or 3.62) the mean radius of our configuration which we shall hereafter (there should be no danger of confusion) denote by r. On the other hand, the functions a^2 and A as defined by Eqs. (4.13) and (4.30) depend on the structure of the envelope surrounding the central mass-point, the oscillations of which we are considering; and our investigation can obviously proceed no further without making specific assumptions about them.

In order to do so, let us assume the oscillating configuration to consist indeed of a central mass-point of mass m, surrounded by an evanescent envelope the density of which, though infinitesimal (so that its gravitational effects can be ignored), is sufficient to render the material opaque and falls off with increasing radius r in accordance with a law $r^{-\alpha}$ (of many that could be considered), where α represents a constant.

In more specific terms, let our model under consideration consist of a discrete central particle of mass

$$\frac{4}{3}\pi R^3\bar{\rho}\left(1 - \frac{3\epsilon}{(3-\alpha)R^\alpha\bar{\rho}}\right),$$

where R denotes the external radius of our configuration, $\bar{\rho}$ its mean density, and ϵ the constant of the density-distribution law of the form

$$\rho = \epsilon r^{-\alpha}, \tag{4.66}$$

valid for $0 < r$, such that ϵ can be made to approach zero.

The mass $m(r)$ of such a configuration inside a sphere of radius r will obviously be given by

$$m(r) = \frac{4}{3}\pi R^3\bar{\rho}\left(1 - \frac{3\epsilon}{(3-\alpha)R^\alpha\bar{\rho}}\right) + 4\pi\epsilon\int_0^r r^{2-\alpha}\,dr, \tag{4.67}$$

leading to a total mass of

$$m(R) = \tfrac{4}{3}\pi R^3\bar{\rho} \tag{4.68}$$

and the gravitational acceleration

$$g = \frac{Gm(r)}{r^2} = \frac{4\pi G R^3\bar{\rho}}{3r^2}\left(1 - \frac{3\epsilon}{(3-\alpha)R^\alpha\bar{\rho}}\right) + \frac{4\pi G\epsilon}{3-\alpha}r^{1-\alpha}, \tag{4.69}$$

where G denotes, as before, the gravitational constant. Moreover, by an integration of the equation

$$dP/dr = -g\rho \tag{4.70}$$

of hydrostatic equilibrium, we find from (4.69) and (4.70) that

$$P = \frac{4\pi G R^3 \bar{\rho}\epsilon}{3(\alpha+1)}\left(1 - \frac{3\epsilon}{(3-\alpha)R^\alpha \bar{\rho}}\right)\left(\frac{1}{r^{\alpha+1}} - \frac{1}{R^{\alpha+1}}\right) + \frac{2\pi G\epsilon^2}{(3-\alpha)(\alpha-1)}\left\{\frac{1}{r^{2(\alpha-1)}} - \frac{1}{R^{2(\alpha-1)}}\right\}. \tag{4.71}$$

As $\epsilon \to 0$, both the pressure P and density ρ tend to zero (except at the center), but in such a way that their ratio remains finite, and equal to

$$\frac{P}{\rho} = \frac{4\pi G R^2 \bar{\rho}}{3(\alpha+1)}\left(\frac{1 - x^{\alpha+1}}{x}\right), \tag{4.72}$$

where we have abbreviated

$$x = r/R. \tag{4.73}$$

Accordingly, the square of the local velocity of sound

$$a^2 = \gamma\frac{P}{\rho} = \frac{4\pi G R^2 \bar{\rho}\gamma}{3(\alpha+1)}\left(\frac{1 - x^{\alpha+1}}{x}\right), \tag{4.74}$$

and

$$A = \frac{1}{Rx}\left(\alpha - \frac{\alpha+1}{\gamma(1 - x^{\alpha+1})}\right). \tag{4.75}$$

If we now introduce (4.74) and (4.75) together with (4.73) in (4.64), the latter can be readily shown to assume the form

$$(1 - x^{\alpha+1})\, x^2 \frac{\partial^2 u_\xi}{\partial x^2} + (1 - \alpha - 2x^{\alpha+1})x\frac{\partial u_\xi}{\partial x} + \left(\frac{2(\alpha+1)(2-\gamma)}{\gamma} - 2(1 - x^{\alpha+1}) + \tilde{\omega}^2 x^3\right) u_\xi = 0, \tag{4.76}$$

where we have abbreviated

$$\tilde{\omega}^2 = \frac{3(1+\alpha)}{4\pi G\bar{\rho}\gamma}\,\tilde{\nu}^2, \tag{4.77}$$

and which for the particular case of $\alpha = 2$ reduces further to

$$(1 - x^3)\, x^2 \frac{\partial^2 u_\xi}{\partial x^2} - (1 + 2x^3)x\frac{\partial u_\xi}{\partial x} + \left[4\left(\frac{3}{\gamma} - 2\right) + (\omega^2 + 2)\, x^3\right] u_\xi = 0. \tag{4.78}$$

On substituting

$$x^3 = y \tag{4.79}$$

this equation transforms to

$$9(1-y)y^2\frac{\partial^2 u_\xi}{\partial y^2}+3(1-4y)y\frac{\partial u_\xi}{\partial y}+\left[4\left(\frac{3}{\gamma}-2\right)+(\omega^2+2)y\right]u_\xi=0. \tag{4.80}$$

The latter can be recognized (cf. Kopal, 1950a) as a hypergeometric equation, the solution of which can be expressed in the form of a hypergeometric series

$$\begin{aligned} u_\xi &= B_{1,2}\,y^k F(a,b,c;y),\\ &= B_{1,2}x^{\pm(J+i)}F(a,b,c;x^3), \end{aligned} \tag{4.81}$$

where $B_{1,2}$ are arbitrary integration constants, but

$$k=\pm\tfrac{1}{3}(J+1) \tag{4.82}$$

and

$$a,b=\frac{3+2J\pm\sqrt{9+4\omega^2}}{6}, \tag{4.83}$$

$$c=\frac{3+2J}{3}, \tag{4.84}$$

where

$$J^2=3\left(3-\frac{4}{\gamma}\right). \tag{4.85}$$

Barring the exceptional case when $c = 1$ (or any other positive integer[6]) the series on the right-hand side of Eqs. (4.81) represents two independent solutions of (4.78) or (4.80) and their linear combination furnishes, therefore, its complete primitive.

In order to reduce this primitive to the form in which it can represent the solution of our vibrational problem, appropriate boundary conditions must be brought to bear upon it—imposing, in fact, no other requirements than that this solution be bounded throughout the entire interval of its validity including its end points $x = 0$ and 1. A requirement that the solution remain finite at the origin requires that the integration constant $B_2 = 0$ annihilate the solution corresponding to the negative value of k; while the requirement that u_ξ be finite at $x = 1$ calls for the hypergeometric series $F(a, b, c; 1)$ on the right-hand side of (4.81) to remain convergent. Since, however, $a + b = c$, the series is known to diverge there, and the sum of its terms can remain finite only if we

[6] In such a case, the second independent solution of (4.80) would possess a logarithmic singularity at the origin, and must be removed by the adoption of a suitable integration constant.

terminate the series by setting a or b equal to zero or a negative integer (say, $-i$), which will be the case if, and only if,

$$\tilde{\omega}^2 = (3i + J)(3i + J + 3), \qquad i = 0, 1, 2, 3, \ldots . \tag{4.86}$$

This condition restricts the permissible modes of free oscillations of our model to those with a *discrete* frequency spectrum given by (4.86); and the corresponding eigenfunctions can be shown to assume the form of the Jacobi polynomials

$$u_\xi(x) = B_1 x^{J+1} G_i(c, c, x^3), \tag{4.87}$$

where the constants c and J continue to be given by Eqs. (4.84) and (4.85), and B_1 represents an arbitrary scale parameter. The discrete nature of the frequency spectrum of this problem was first recognized by Sterne (1937), and its hypergeometric character by Kopal (1950a).

On the other hand, it can be shown (cf. Prasad, 1953; or Simon, 1958) that if we set out to expand the solution of Eq. (4.80) in ascending powers of $1 - y$ rather than y, this equation admits also of a particular solution of the form

$$u_\xi(x) = B_1 x^{J+1} F(a, b, 1; 1 - x^3), \tag{4.88}$$

where a, b, and J continue to be given by Eqs. (4.83) and (4.85). Since, in the present case, $c = 1$, the second independent solution of (4.80), factored by the integration constant B_2, would become infinite at $x = 1$ on account of its logarithmic singularity and must be removed by setting $B_2 = 0$; while, obviously, $B_1 = u_\xi(1)$. Since, however, with the *negative* root of (4.85) for J,

$$a + b = 1 + \tfrac{2}{3}J < 1, \tag{4.89}$$

and the hypergeometric series on the right-hand side of (4.88) then converges also for $x = 0$ regardless of the frequency $\tilde{\nu}$, and if, in addition, $J \geqslant -1$, the particular solution (4.88) of Eq. (4.78) remains regular at the origin.

Therefore, for any root of (4.85) constrained by the inequality

$$-1 \leqslant J < 0, \tag{4.90}$$

Eq. (4.88) represents another eigensolution of our problem—analogous to the polynomial solutions of the form (4.87). But, unlike the latter, its existence does not restrict in any way the frequency in which the configuration can oscillate radially. In other words, the frequency spectrum associated with (4.88) is not discrete (as it proved to be in the former case) but *continuous* over the entire domain of $0 < \tilde{\nu}^2 < \infty$.

What distinguishes the two types of eigenfunctions associated with the

continuous and discrete spectra is the variation of u_ξ with x in the proximity of the origin: for sufficiently small values of x they both vary as x^{J+1}, but their frequency spectra will be continuous or discrete depending on whether $J \gtrless 0$. Discrete spectrum can embrace terms for which J is positive (or, if negative, $J \geqslant -1$), but continuous spectrum can exist only if $0 > J \geqslant -1$, which corresponds (by Eq. 4.85) to a range in γ given by the inequality

$$\tfrac{4}{3} \leqslant \gamma \leqslant \tfrac{3}{2}. \tag{4.91}$$

For $\gamma = 1.5$, the continuous spectrum associated with (4.88) disappears from the real domain, and only discrete eigenvalues (4.86) corresponding to (4.87) are admissible.

Should $J \geqslant 0$, the condition (4.89) is not satisfied and the series on the right-hand side of (4.88) diverges. It can remain bounded only if the infinite series is truncated to polynomial form by setting a or b equal to $-i$ ($i = 0, 1, 2, 3,...$); but in doing so we recover merely the discrete spectrum (4.86) and the polynomials (4.87).[7]

The results obtained above represent the *radial* oscillations which our particular Roche model is capable of in the proximity of the origin; and as such, they can be regarded as the inner boundary condition of the more general problem considered in Section III,B in the immediate proximity of the origin. On the other hand, the outer boundary conditions continue to be met by the requirement that the oscillations remain finite on the outer surface $\rho(\xi_1) = 0$. While this latter condition is indeed necessary for the stability of our configuration as a whole, the requirement that oscillations be radial near the origin is not unique. Such oscillations can be also nonradial—characterized, for example, by spheroidal rather than spherical symmetry. The possibility of such nonradial oscillations has indeed also been studied before (cf. Ottelet, 1960; Kopal, 1968), but only in the very restricted case of $\alpha = 2$ and $\gamma = 1.5$, and even then the analytical structure of the problem becomes so complicated that for fuller details we must refer the reader to the original publications.

Lastly, it should be reiterated that all developments contained in this section hold good only in the immediate neighborhood of the origin, where our configuration can retain spherical symmetry. As soon as this ceases to be true with increasing distance from the center, and the right-hand sides of Eqs. (4.65) for the metric coefficient h_1 and its logarithmic derivative must be augmented by terms arising from the distortion, it can be shown (cf. Kopal, 1970; section 5) that it is no longer possible to satisfy Eqs. (4.32)–(4.34) by regarding u_ξ as a function of ξ only, and disregarding u_η or u_ζ. In such a case we are faced with

[7] If the condition (4.89) is not met for nonnegative values of J, and the series $F(a, b, 1; 1 - x^3)$ then diverges at the origin, the right-hand side of Eq. (4.98) becomes an indeterminate product, but such that $\lim_{x\to 0} x^{J+1}F(a, b, 1; 1 - x^3) = \infty$.

a necessity to solve for them from simultaneous systems of *partial* differential equations of the form (4.32)–(4.34), but this represents a task for the future.

D. *Generalized Roche Model*

In concluding our discussion of the Roche model as given in this chapter, we wish to point out some aspects of the stability of a generalized Roche model, differing from the one discussed in the preceding sections in so far as its central mass is not confined to a mass-point at its center, but occupies a sphere of finite radius. In so far as this mass retains spherical symmetry, the potential arising from it will still be represented by the first term on the right-hand sides of Eqs. (2.1) or (2.5) regardless of the finite size of the central core; but its finite dimensions will, on the other hand, vitally affect the vibrational stability of such a composite model. In the present section, an outline of the stability of such a model with respect to free radial oscillations will be given.

The concept of a "generalized Roche model" consisting of a massive core of finite dimensions, which is surrounded by an evanescent envelope whose contribution to the gravitational potential of the composite configuration can be ignored, was introduced in the astronomical literature by Jeans (1919), in connection with his work on the origin of the solar system. Subsequent work on the vibrational stability of this composite model included contributions by Sen (1943), Kopal (1950b), Prasad (1953), Simon (1958), and Gurm (1960); and the aim of the present section will be to outline the principal results obtained in the field so far.

To begin with, consider a spherical model undistorted by rotation or tides, and consisting of a fluid core of finite fractional dimensions surrounded by an evanescent atmosphere. If the fluid constituting the core were *incompressible*, the core would be incapable of performing purely radial oscillations; so that within the scheme of our approximation we must consider it to be at rest. The infinitesimally thin atmosphere surrounding it can oscillate freely around it provided that the velocity of the oscillation vanishes at the interface between the core and the envelope, i.e., provided this interface becomes a model surface of oscillating motion; and as long as the core is regarded as incompressible, the pressure variation at the interface is clearly irrelevant. Moreover, if the atmospheric density falls off with the inverse α-th power of radial distance in accordance with Eq. (4.66), the radial velocity u_ξ of oscillatory motion continues to be given by Eq. (4.76) which, for the particular case of $\alpha = 2$, reduces to (4.78) or (4.80).

Their solutions of the form (4.81) or (4.88) continue, therefore, to apply from the boundary $x = 1$ down to $x = x_c$ representing the fractional radius of the core. However, the fact that, by definition of the generalized Roche model, $x_c > 0$ makes it unnecessary to set $B_2 = 0$ in (4.81) to keep u_ξ bounded within

the domain of its validity; nor do we have to be concerned about the divergence of the series on the right-hand side of (4.88) for $x = 0$. What we do have to ensure is that

$$u_\xi(x_c) = 0 \tag{4.92}$$

at the interface, which means that x_c must be identical with one of the zeros of the respective eigenfunctions representing u_ξ. For polynomial eigenfunctions of the form (4.87) their zeros for $i = 1, 2, 3$, and $J = 0, \pm\sqrt{0.6}, \pm\sqrt{1.2}, \pm\sqrt{1.8}$ have been tabulated to 6D by Kopal (1950b, Table 1). On the other hand, the eigenfunctions of the form (4.88) possess an infinite number of zeros, of which those comprising the closed interval $(x_c, 1)$ will correspond to higher modes of oscillation.

In the next model which we wish to consider, the core will be treated as consisting of a *compressible* fluid, capable of free radial oscillations of its own. Suppose, for the sake of argument, that the envelope surrounding it has been excited to oscillate in the ith mode. Is it possible that it can continue to do so without disturbing the core? The answer would be in the affirmative only provided that a sphere of no displacement could be found at which the pressure remains constant, and that this sphere encloses the inner core. It is, however, easy to prove that this requirement cannot be met; for whereas no motion will occur if $u_\xi(x_c) = 0$, the constancy of pressure would require that also the derivative $\partial u_\xi/\partial x$ vanish at the same point—which (by Rolle's theorem) is impossible for a function which is not identically zero. In oscillating configurations, the zeros of pressure and velocity variations cannot coincide anywhere in the interior. Oscillations in the course of which the pressure varies at the interface are, indeed, bound to excite synchronous oscillations in the compressible core if the latter is capable of them; otherwise they will get dissipated.

And this leads us to consider the third, and most general, case in which the entire composite configuration can oscillate as a whole. The first and obvious condition for the possibility of coexistence of such a regime is a requirement that the respective free periods of radial oscillation of the core and its envelope be equal. Consistent with the assumption underlying the Roche model that the mass of the envelope is infinitesimal in comparison with that of the core, let us consider first that free oscillations of the core are unaffected by its surrounding envelope. Accordingly, the frequency of oscillations of the configuration as a whole should be controlled by those of the core; and the envelope floating on this core must adjust to these if it can. Our immediate problem reduces thus to an investigation of the conditions under which this may indeed be the case.

In order to do so, we must first inquire about the frequency spectrum of free oscillations of the core of radius r_c/R. If, for the sake of simplicity, we regard

this core as a homogeneous globe of constant density ρ_c, it can be shown that, inside the core,

$$a^2 \equiv \gamma P_0/\rho_0 = \tfrac{2}{3}\pi G\rho_c r_c^2 \gamma(1 - x^2) \tag{4.93}$$

and

$$A = -2r_c x^3/\gamma(1 - x^2), \tag{4.94}$$

where x now stands for the ratio r/r_c. When (4.93) and (4.94) are inserted in (4.64),[8] the result will be the equation

$$(1 - x^2)x^2 \frac{\partial^2 u_\xi}{\partial x^2} + 2(1 - 2x^2)x \frac{\partial u_\xi}{\partial x} + [(4 + \tilde{\omega}_c^2 - \tfrac{2}{3}J^2)x^2 - 2]\, u_\xi = 0, \tag{4.95}$$

where J continues to be given by (4.85) and

$$\tilde{\omega}_c^2 = 3\tilde{\nu}_c^2/2\pi G\rho_c\gamma. \tag{4.96}$$

The present writer has shown elsewhere (cf. Kopal, 1948) that a substitution $x^2 = y$ transforms (4.95) again to a hypergeometric equation, and that a requirement of regularity at the origin restricts its particular solution to be of the form of a hypergeometric series

$$u_\xi(x) = AxF(a, b, a + b, x^2), \tag{4.97}$$

where

$$a, b = \frac{5 \pm (25 + 4\tilde{\omega}_c^2 - \tfrac{8}{3}J^2)^{1/2}}{4} \tag{4.98}$$

and A is an integration constant.

The second part of our boundary conditions, requiring u_ξ to remain finite on the surface, necessitates that the hypergeometric series on the right-hand side of (4.97) either converge for $x = 1$ or, if divergent, must be terminating. Since, however, in the present case the sum of the first two indices of our series equals the third, the hypergeometric series is known to diverge over unit radius, and must be reduced to a polynomial if $u_\xi(1)$ is to remain bounded. Doing so by equating a or b as given by (4.98) to zero or a negative integer $-i$, the admissible values of $\tilde{\omega}_c$ are thereby restricted to

$$\tilde{\omega}_c^2 = (2i + 5)\, i + \tfrac{1}{3}J^2, \qquad i = 0, 1, 2, \ldots, \tag{4.99}$$

[8] Care merely being taken to note that, in a homogenous core, the symbol k in Eq. (4.21), which previously we normalized out to unity, will now be equal to

$$k = \frac{1}{\rho_0}\frac{\partial P_0}{\partial \xi} = \frac{4}{3}\pi G\rho_c r_c^4 x^3,$$

and its normalized value becomes equal to $2x^3$ in place of 1.

which, inserted in Eq. (4.97), reduces the eigenfunctions of our present problem to

$$u_\xi(x) = AxG_i(\tfrac{5}{2}, \tfrac{5}{2}, x^2), \tag{4.100}$$

where G_i denotes the respective Jacobi polynomial.

Equation (4.99) combined with (4.96) leads, furthermore, to

$$\tilde{\nu}_c^2 = \tfrac{4}{3}\pi G\rho_c\gamma[(2i + 5)\, i + \tfrac{1}{3}J^2] \tag{4.101}$$

as the set of discrete frequencies in which a homogeneous compressible sphere, disturbed slightly from its state of equilibrium, may perform radial oscillations. These frequencies are, moreover, real provided that the ratio γ of specific heat of the material satisfies the inequality

$$\gamma > 4/3 + (2i + 5)i; \tag{4.102}$$

the limit of γ below which the instability sets in depends, therefore, on the degree of the excited eigenfunction. This discrete nature of the characteristic frequency spectrum of a homogeneous compressible gas sphere has likewise been first established by Sterne (1937), and its hypergeometric nature by Kopal (1948).

Returning now to our composite configuration in which a homogeneous core is surrounded by an evanescent envelope in which the density falls off with inverse square of the radial distance, we readily see that such a configuration can oscillate radially as a whole only if the core and its envelope can do so in unison, i.e., in such frequencies for which their characteristic spectra

$$\tfrac{4}{3}\pi G\bar{\rho}\gamma(3i + J)(3i + J + 3)$$

and

$$\tfrac{4}{3}\pi G\rho_c\gamma[(2i + 5)\, i + \tfrac{1}{3}J^2]$$

overlap (with $\bar{\rho}/\rho_c = r_c^3/R^3$, i being not necessarily the same in both). The restrictions arising from such a requirement are severe and rather complicated (for their details, see Kopal, 1950b), but their complexity goes wholly back to our initial assumption that the interface between the core and its envelope as well as the external radius of our configuration constitute nodal surfaces of our oscillation problem; which follows, in turn, from our assumption that the evanescent envelope produces no pressure at the base of the atmosphere.

If this assumption is removed, however (cf. Prasad, 1953; Simon, 1958), the problem becomes considerably simpler: for if the envelope exerts a *finite* pressure on the core, the validity of the solution (4.97) needs to extend only for $x < 1$ to such a point at which the residual pressure in the core (varying as $1 - x^2$) can be balanced by that of the atmosphere; and for $x < 1$ the hypergeometric series on the right-hand side of (4.97) is known to converge for *any*

values of the parameters a or b. In other words, *a homogeneous compressible core subjected to uniform outside pressure at the boundary can oscillate in any frequency*; and, moreover, for $x > x_c > 0$ the eigensolution (4.90) for the envelope can do the same as well. Therefore, radial oscillations of a generalized Roche model envisaged in the latter part of this section are not restricted to any set of discrete frequencies; and the spectrum of the latter becomes continuous.

It is true that by allowing the envelope to exert a finite pressure on a core of finite distension we have departed somewhat from an orthodox definition of the Roche model. However, this departure from consistency may be more than made up by its greater physical reasonableness; and on the basis of physical consideration we should, therefore, expect that a generalized Roche model consisting of compressible core as well as envelope should be able to oscillate radially with any arbitrary frequency.

In conclusion, we wish to note the investigations of the form of static generalized Roche model with nonhomogeneous distorted cores initiated by Takeda (1934), and resumed later by Roxburgh *et al.* (1965).

APPENDIX

In Section II,A of this chapter we introduced the Roche equipotential surfaces Ω or ξ in terms of the rectangular or spherical polar coordinates x, y, z or r, λ, ν. It may, however, be of interest to note certain coordinate transformations—real or complex—which may be applied to render the form of the Roche potentials more symmetrical.

Thus, introducing a complex transformation

$$\begin{aligned} u^2 &= x + i(y^2 + z^2)^{1/2} \\ s^2 &= x - i(y^2 + z^2)^{1/2} \\ t^2 &= z^2 u^{-2} \end{aligned} \tag{A.1}$$

or its inverse

$$\begin{aligned} x^2 &= +\tfrac{1}{4}(u^2 + s^2)^2 \\ y^2 &= -\tfrac{1}{4}(u^2 - s^2)^2 - u^2 t^2 \\ z^2 &= +u^2 t^2 \end{aligned} \tag{A.2}$$

in place of (2.4), we find that

$$\begin{aligned} r &= us, \\ r' &= \sqrt{(1 - u^2)(1 - s^2)}, \end{aligned} \tag{A.3}$$

and

$$\lambda = (u^2 + s^2)/2us, \qquad \nu = t/s. \tag{A.4}$$

Accordingly, the expression (2.5) for the potential ξ assumes the form

$$\xi = \frac{1}{us} + q\left(\frac{1}{[(1-u^2)(1-s^2)]^{1/2}} - \frac{u^2+s^2}{2}\right) + \frac{q+1}{2}u^2(s^2-t^2); \quad \text{(A.5)}$$

or, in trigonometric terms,

$$\begin{aligned} \xi + \tfrac{1}{2}q = {} & \csc\alpha \csc\beta + \tfrac{1}{2}q[2\sec\alpha\sec\beta + \cos(\alpha+\beta)\cos(\alpha-\beta)] \\ & + \tfrac{1}{2}(q+1)\sin^2\alpha \sin(\beta+\gamma)\sin(\beta-\gamma), \end{aligned} \quad \text{(A.6)}$$

where we have abbreviated

$$\begin{aligned} u &= \sin\alpha, \\ s &= \sin\beta, \\ t &= \sin\gamma. \end{aligned} \quad \text{(A.7)}$$

Moreover, by use of hyperbolic functions

$$\begin{aligned} u^2 + s^2 &= 2\cosh E \cosh F, \\ u^2 - s^2 &= -2\sinh E \sinh F, \end{aligned} \quad \text{(A.8)}$$

we find that

$$r' = \sqrt{(1-u^2)(1-s^2)} = \cosh F - \cosh E, \quad \text{(A.9)}$$

but

$$r = us = \sqrt{\sinh^2 E + \cosh^2 F}. \quad \text{(A.10)}$$

On the other hand, a real transformation

$$\begin{aligned} 1 - x &= p^2 - q^2 - s^2 \\ y &= 2pq \\ z &= 2ps \end{aligned} \quad \text{(A.11)}$$

permits us to write

$$r' = p^2 + q^2 + s^2; \quad \text{(A.12)}$$

but, unfortunately, the simplification results only at the price of more complicated expressions for r, λ, and ν.

The transformation (A.1) in the plane $z = 0$ was introduced by Hill (1878); and that represented (likewise for $z = 0$) by (A.11) is due to Levi-Civita (1906)—both in connection with their work on the Jacobian integral of the restricted problem of three bodies. A generalization of such coordinates to three dimensions was later given by Kopal (1959; sec III.1).

LIST OF SYMBOLS

Symbol	Meaning
m, m'	Masses of the primary (m) and secondary (m') components in a close binary system
q	Mass-ratio m'/m (Eq. 2.7)
ω	Angular velocity of axial rotation of mass m
$2n$	Normalized value of ω^2 (Eq. 2.9)
Ω	Total potential of forces acting on mass m (Eq. 2.1)
ξ	Normalized value of the potential Ω, as defined by Eq. (2.5); also one of the Roche curvilinear coordinates
r, r'	Distance of point P (cf. Fig. 2) from the center of masses m and m'
r_0	Mean fractional radius of an equipotential surface ξ = constant
V	Volume of an equipotential ξ = constant (cf. Eq. 2.19)
r_c	Radius of the core of a generalized Roche model
x, y, z	Cartesian rectangular coordinates of point P on Fig. 2
ν, θ, ϕ	Spherical polar coordinates of an arbitrary point
λ, μ, ν	Direction cosines of an arbitrary point (cf. Eqs. 2.4)
ξ, η, ζ	Roche curvilinear coordinates
$h_{1,2,3}$	Metric coefficients of the Roche transformation (3.4)
l_j , m_j , n_j	Direction cosines of the normal, tangent, and binormal to the equipotential surfaces (as defined by Eqs. 3.8–3.10)
$u_{\xi,\eta,\zeta}$	Velocity components in the direction of increasing Roche coordinates ξ, η, ζ
Δ	Divergence of the velocity vector in curvilinear coordinates (Eq. 4.11)
∇_1	Laplacian operator in curvilinear coordinates (Eq. 4.45)
P	Pressure at any point in the interior
ρ	Density at any internal point
$\bar{\rho}$	Mean density of the Roche model
ρ_c	Mean density of the core of a generalized Roche model
a	Velocity of sound (Eq. 4.13)
γ	Ratio of specific heats
R	Mean radius of the Roche model
g	Gravitational acceleration
G	Gravitation constant
$\tilde{\nu}$	Frequency of free oscillations (Eq. 4.31)
$\tilde{\omega}$	Normalized frequency of free oscillations
$P_j(x)$	Legendre polynomials (zonal harmonics) of argument x and degree j
$F(a, b, c; x)$	Ordinary hypergeometric series of the type ${}_2F_1$
$G_i(p, q, x)$	Jacobi polynomials in x of degree i
$R_j(\lambda_0 , \nu_0)$	Roche harmonics (Eq. 4.59)
$L_{1,2,3}$	Lagrangian collinear points in the restricted problem of three bodies
$L_{4,5}$	Lagrangian triangular points in the same problem
$C_{1,\dots,5}$	Values of the Jacobi constants assumed by the potential at the Lagrangian points $L_{1,\dots,5}$, and defined by the Eqs. (2.38), (2.88), and (2.93), respectively.

References

Chandrasekhar, S. (1933). *Mon. Notic. Roy. Astron. Soc.* **93**, 390–405, 449–461, 462–471, and 539–574.

Darboux, G. (1910). "Leçons sur les systèmes orthogonaux et les coordonnées curvilignes," Chapter II. Gauthier-Villars, Paris.
Forsyth, A. R. (1912). "Lectures on Differential Geometry of Curves and Surfaces." Cambridge Univ. Press, London and New York.
Gurm, H. S. (1960). *Proc. Nat. Inst. Sci. India* **26**, 480–485.
Hill, G. W. (1878). *Amer. J. Math.* **1**, No. 5, 129 and 245.
Jeans, J. H. (1919). "Problems of Cosmogony and Stellar Dynamics," Cambridge Univ. Press, London and New York. Sect. 154.
Kitamura, M. (1970). *Astrophys. Space Sci.* **7**, 272–358.
Kopal, Z. (1948). *Proc. Nat. Acad. Sci. U. S.* **34**, 377–384.
Kopal, Z. (1950a). *Proc. Nat. Acad. Sci. U. S.* **36**, 72–84.
Kopal, Z. (1950b). *Astrophys. J.* **111**, 395–407.
Kopal, Z. (1954a). *Mem. Soc. Roy. Sci. Liege* **15**, 684.
Kopal, Z. (1954b). *Jodrell Bank Ann.* **1**, 37–57.
Kopal, Z. (1956). *Ann. Astrophys.* **19**, 298–335.
Kopal, Z. (1959). "Close Binary Systems," Chapter III. Wiley, New York.
Kopal, Z. (1960). "Figures of Equilibrium of Celestial Bodies." Univ. of Wisconsin Press, Madison.
Kopal, Z. (1968). *Astrophys. Space Sci.* **2**, 48–60.
Kopal, Z. (1969). *Astrophys. Space Sci.* **5**, 360–384.
Kopal, Z. (1970). *Astrophys. Space Sci.* **8**, 149–171.
Kopal, Z. (1971). *Astrophys. Space Sci.* **10**, 328–331.
Kopal, Z., and Ali, A. K. M. S. (1971). *Astrophys. Space Sci.* **11**, 423–429.
Kopal, Z., and Kitamura, M. (1968). *Advan. Astron. Astrophys.* **6**, 125–172.
Kuiper, G. P., and Johnson, J. R. (1956). *Astrophys. J.* **123**, 90–94.
Lamb, H. (1932). "Hydrodynamics," 6th ed., pp. 156–159. Cambridge Univ. Press, London and New York.
Levi-Civita, T. (1906). *Acta Math.* **30**, 305–327.
Lichtenstein, L. (1928). *Ber. Sachs. Akad. Wiss.* **80**, 35–68.
Maruhn, K. (1931). *Math. Z.* **33**, 300–320.
Ottelet, I. (1960). *Ann. Astrophys.* **23**, 218–231.
Plavec, M., and Kratochvil, P. (1964). *Bull. Astron. Inst. Czech.* **15**, 165–170.
Poincaré, H. (1892). "Les méthodes nouvelles de la mécanique céleste," Vol. 1, Chapter V. Gauthier-Villars, Paris.
Prasad, C. (1953). *Proc. Nat. Inst. Sci. India* **19**, 739–745.
Prendergast, K. H. (1960). *Astrophys. J.* **132**, 162–174.
Roche, E. A. (1849). *Mem. Acad. Sci. Montpellier* **1**, 243 and 333.
Roche, E. A. (1851). *Mem. Acad. Sci. Montpellier* **2**, 21.
Roche, E. A. (1873). *Ann. Acad. Sci. Montpellier* **8**, 235.
Rosenthal, J. E. (1931). *Astron. Nachr.* **244**, 169.
Roxburgh, I. W., Griffith, J. S., and Sweet, P. A. (1965). *Z. Astrophys.* **61**, 203–221.
Sen, H. K. (1943). *Proc. Nat. Acad. Sci. India* **13**, 44–53.
Simon, R. (1958). *Astrophys. J.* **127**, 428–435.
Sterne, T. E. (1937). *Mon. Notic. Roy. Astron. Soc.* **97**, 582–593.
Szebehely, V. G. (1967). "Theory of Orbits: The Restricted Problem of Three Bodies," Chapters 4 and 10. Academic Press, New York.
Takeda, S. (1934). *Mem. Coll. Sci., Kyoto Imp. Univ., Ser. A* **17**, 197–217.
ten Bruggencate, P. (1934). *Z. Astrophys.* **8**, 344–357.
Tisserand, F. (1883). *C. R. Acad. Sci.* **96**, 1914.

Lunar Eclipses

FRANTIŠEK LINK*
Astronomical Institute of the Czechoslovak Academy of Sciences
Prague, Czechoslovakia

* Present address: Centre National de la Recherche Scientifique, Institut d'Astrophysique, Paris.

I

INTRODUCTION

From the earliest beginnings of civilization, the picturesque phenomenon of lunar eclipses, unrivaled but for those of the Sun, has continued to challenge man's speculative spirit. We speak of a lunar eclipse when the Earth's shadow is cast over the lightless Moon, which causes its temporary obscuration observable all over the terrestrial hemisphere facing it.

An analysis of this definition showed, however, essential *deviations* from the purely geometrical conditions, and revealed numerous new facts concerning the Earth's atmosphere and the Moon itself. This chapter is introduced by a short historical review featuring the part played by lunar eclipses in the evolution of astronomy; the classical theory of lunar eclipses will then be outlined, dealing with the geometrical aspects of the eclipse, together with the determination of the radius of the Earth's shadow and of its enlargement. It will be in this section that the first deviations from geometrical theory will make their appearance. These deviations will take us to the photometric theory of lunar eclipses, the origin of which may be found in the 1930's: in this theory a prominent role is played by the Earth's atmosphere. Further on, an account will be given of the methods of photometric observations of eclipses, and the results of these observations will be compared with the theory. From the discrepancies between theory and observations, new notions concerning the terrestrial atmosphere will arise. Lunar eclipses, moreover, led to the discovery of the phenomenon of lunar luminescence, particularly prominent in the penumbra. Special attention will be given to some of these results, and to further potentialities of these phenomena.

Several sections of this present chapter will also deal with the problems of other phenomena related in one way or another to lunar eclipses. This analysis has significantly contributed to a better understanding of the Moon and other celestial bodies.

II

THE HISTORY OF LUNAR ECLIPSES

Reports on lunar eclipses are available from the most remote past. Schoch in 1931 published records on eclipses observed in Mesopotamia as early as 2202 and 2282 B.C. From China we have another such record dating back to 1135 B.C. (Gaubil, 1740). Later on, the number of reports on lunar eclipses increased. From the beginning of the 8th century B.C., we have further records from

Mesopotamia, Greece, and from other parts of the Mediterranean region, supplemented also by reports from the rest of Europe. Lists of eclipses observed in ancient times, together with some circumstances concerning the phenomenon itself, were assembled by Calvisius (1650), Riccioli (1651), Struyck (1740), and, from the Chinese region, by Gaubil (1740). All these reports are excerpts from chronicles or similar historical documents. From the last half of the 17th century, reports of observations of lunar eclipses began to be published in transactions of various scientific societies arising at that time. Bibliographically, until the end of the 18th century, these works were registered by Reuss (1804). An extensive source of records on observations from the 17th century is also the compilation by Pingré-Bigourdan (1901).

In the 19th century, we find observations of lunar eclipses scattered all over various astronomical and general scientific magazines. Various bibliographical expedients are available from this period. A good guide in looking up eclipses is Oppolzer's detailed catalog (1887) of computed eclipses from 1200 B.C. until until A.D. 2163. The history of eclipses is briefly dealt with by Chambers (1899).

Lunar eclipses also constitute an important chronological expedient, since various historical events may be localized in time according to them. They have been frequently used from ancient times on, and the lists of observed eclipses served this purpose. For instance, Newton made use of three eclipses reported by Ptolemy (namely, those of B.C. 523, VII 16; B.C. 502, XI 19; and B.C. 491, IV 25) to fix the *terminus a quo* from which accordingly can be dated the prophecy of Daniel, the restoration of the Jews under Artaxerxes in 457 B.C., and consequently also the death of Christ.

Lunar eclipses played an important part in the evolution of astronomy. As early as in the 4th century B.C., Aristotle (Prantl, 1881) mentions the circular edge of the terrestrial shadow projected on the Moon as proof of the spherical shape of the Earth. Aristarchos of Samos (Heath, 1913), in the 3rd century B.C., and after him Hipparchos in the 2nd century B.C. (Wolf, 1877a), suggested the use of eclipses for the determination of the system Earth–Moon–Sun. Hipparchos (Wolf, 1877b) also reports on simultaneous observations of lunar eclipses from two localities as a means of establishing the difference of the geographical longitudes of these places. Ptolemaios, in the 2nd century A.D., resorted to ancient and contemporaneous observations of lunar eclipses for a more exact formulation of the theory of lunar motion. This method was also used later on.

In the 17th and 18th centuries, lunar eclipses were beginning to be used for the determination of the differences of geographical longitudes, particularly when—as suggested by Langrenus (1644)—the contacts of craters with the shadow were observed, which fact palpably increased the accuracy of the method.

Later on this method, attaining an accuracy of 0.1 min at the utmost, was replaced by other more exact methods. Nevertheless, it rendered good service during its time. Peiresc (Doublet, 1922), for instance, established in 1634 from

observations of eclipses in Cairo, Aleppo, and in Europe, that the Mediterranean Sea was some 1000 km shorter than assumed at that time.

The first to deal with the photometric aspect of eclipses was Kepler (Frisch, 1858), who explained the Moon's illumination in the shadow by a refraction of solar rays in the terrestrial atmosphere. After him, this problem was rather thoroughly studied by Du Séjour (1786), whose work, published under a too general title, long escaped the attention of astronomers dealing with the eclipses.

If we disregard some minor studies by various authors writing during the 19th century, the interest in lunar eclipses became enlivened in connection with the problem of the enlargement of the Earth's shadow. This problem was investigated particularly by Brosinsky (1888) and Hartmann (1891) and in continuation of their works in Hepperger (1895) and Seeliger's theory (1896) from the end of the 19th century. In the 1920's, it was Danjon's results concerning the effects of solar activity on the brightness of the eclipse that aroused attention. In their wake, and also stimulated by the interest in the upper atmosphere, came Fesenkov's, M. de Saussure's, and Link's photometrical theories of eclipses, as will be shown in more detail.

III

THE BASIC CHARACTERISTICS OF THE ECLIPSES

Tangent to the Earth illuminated by the Sun are two conic shadows, one of which, the full shadow or umbra, is convergent, the other, penumbra, divergent. The surfaces of both these shadow cones are tangent to the Sun as well as to the Earth (Fig. 1). Both shadows are projected on a plane perpendicular to the Moon's radius-vector as concentric circles, the angular dimensions of which, as observed from the Earth, are

$$\begin{aligned} \sigma_1 &= \pi_{☾} + \pi_{\odot} + R_{\odot}\,, \quad &&\text{for the penumbra,} \\ \sigma_2 &= \pi_{☾} + \pi_{\odot} - R_{\odot}\,, \quad &&\text{for the umbra.} \end{aligned} \tag{1}$$

If the Moon in its orbit around the Earth passes through this shadow, an eclipse occurs. Thus, eclipses may take place only when the Moon is full, and only when this full Moon stands in close proximity to the node between the lunar orbit and the ecliptic of the Sun. This is the reason why not all full moons are accompanied by eclipses. There are some years which have no eclipses at all (1962, 1966, 1969); other years, on the other hand, will bring one or two; and exceptionally even three eclipses (1982, 2028). The visibility of the eclipse, due to the nature of this phenomenon, is determined solely by the position of the Moon above the horizon of the observer; in all the other respects, its course is
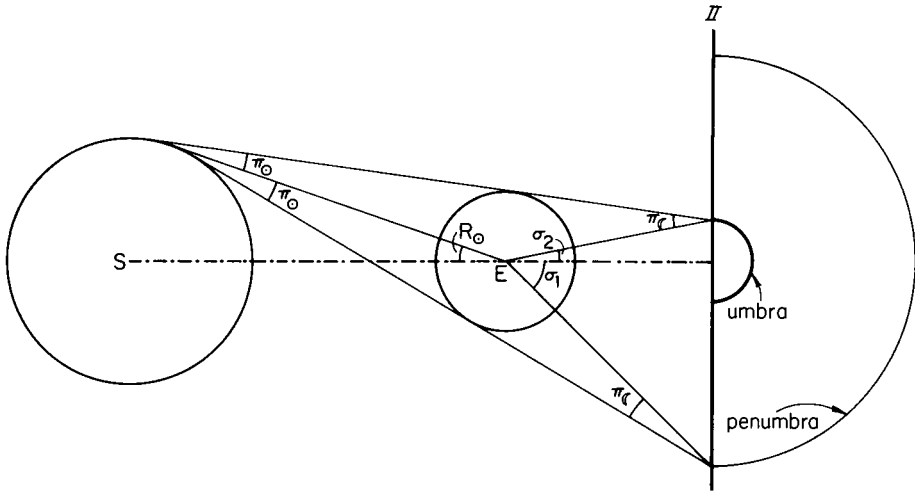
FIG. 1. Umbra and penumbra during lunar eclipses. Half of plane II is rotated into the plane of the picture.

entirely independent of the observer's position. Therefore, even total lunar eclipses are less scarce than the eclipses of the Sun.

The determination of the occurrence of lunar eclipses constitutes one of the elementary problems of spherical astronomy; frequently, a mere graphical solution of this problem will fully satisfy our purpose. Thus, in this respect, it will
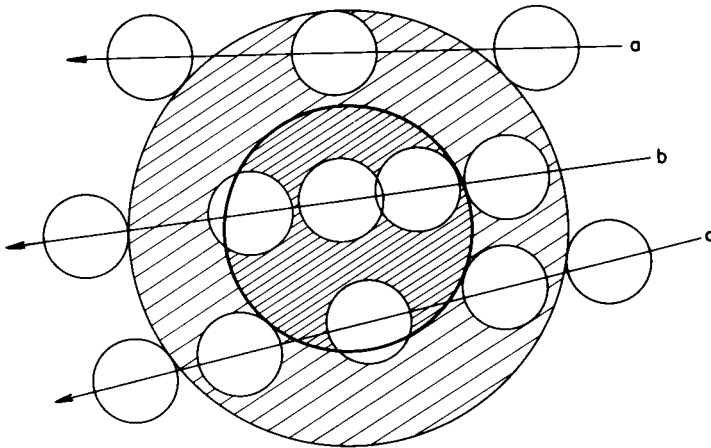
FIG. 2. Diagram of various types of eclipses: (a) penumbral eclipse, (b) total eclipse, (c) partial eclipse.

suffice to refer the reader to some of the standard compendia (Buchanan, 1907; Chauvenet, 1908) on spherical astronomy.

The following brief description of the course of a lunar eclipse will prove useful for later chapters. The beginning of the ingress of the Moon into the penumbra is wholly imperceptible (Fig. 2). Only after the Moon's entrance to about one-half of its diameter into the penumbra may a faint obscuration be observed on its limb. This obscuration then rapidly grows up to first contact of the limb of the Moon with that of the shadow; this marks the commencement of the partial eclipse. At first, the part of the Moon which has entered the shadow is not at all, or only partially, discernible; mainly because of the contrast with the still bright part of the Moon in the penumbra. Only after the greater part of the Moon has entered into the shadow, does its obscured part also begin to become visible again. Sometimes, even, major details on its surface are detectable. The total phase of the eclipse commences with the internal contact of the Moon with the limit of the shadow. At this stage, the Moon, together with the contours of its seas and some of its craters, begins to be discernible, and its color is mostly of a distinct red hue. In the full shadow, the Moon presents a very picturesque spectacle; and its brightness and color, varying from one limb to the other, create the impression of a sphere. After the maximum eclipse, the course of the events is reversed, even if—in contrast to the first phase—the deviations in color and brightness sometimes appear more marked.

IV

FUTURE ECLIPSES

Table I presents a list of future lunar eclipses according to Oppolzer (1887) together with their basic ephemeris. The table contains: (1) The number of the eclipse according to Oppolzer; (2) the Gregorian and (3) the Julian date; (4) the middle of the eclipse in GMT; (5) the magnitude of the eclipse in inches (12 in. equal the Moon's diameter); (6) the semiduration of the partial phase; (7) the semiduration of the total phase; (8) the geographical longitude; and (9) the geographical latitude of the locality, where, at the conjunction, the center of the shadow is in the zenith. Columns (2) and (4) are in accordance with the civilian custom of dating with zero hour at midnight.

Nonnumbered eclipses are penumbral (Bao-Lin, 1964) and their magnitude is given in percentage of the Moon's diameter obscured by the penumbra.

V

INCREASE OF THE EARTH'S SHADOW

As early as at the beginning of the 18th century, astronomers were well aware of the fact that the Earth's shadow projected on the Moon during eclipses,

TABLE I

LIST OF FUTURE UMBRAL AND PENUMBRAL ECLIPSES[a]

(1)	(2)			(3)	(4)	(5)	(6)	(7)	(8)	(9)
					hr min		min	min	deg	deg
4914	1970	II	21	0639	8 31	0.6	26		—124	+11
4915		VIII	17	0816	3 25	5.0	71		—50	—14
4916	1971	II	10	0993	7 42	15.6	107	39	—112	+14
4917		VIII	6	1170	19 44	20.7	112	51	+65	—17
4918		I	30	1347	10 53	12.9	102	21	—160	+18
4919	1972	VII	26	1525	7 18	6.9	80		—108	—20
	1973	I	18	1701	21 18	89%	120			
		VI	15	1849	20 51	49%	106			
		VII	15	1879	11 39	13%	55			
4920		XII	10	2027	1 48	1.2	36		—29	+23
4921	1974	VI	4	2203	22 14	9.9	93		+26	—22
4922		XI	29	2381	15 16	15.5	106	38	+128	+21
4923	1975	V	25	2558	5 46	17.5	109	45	—87	—21
4924		XI	18	2735	22 24	13.1	102	23	+20	+19
4925	1976	V	13	2912	19 50	1.7	43		+62	—18
4926	1977	IV	4	3238	4 21	2.5	51		—64	—6
	1977	IX	27	3414	8 29	93%	131			

[a] From data of Oppolzer (1887) and Bao-Lin (1964).

somewhat exceeded the value given by the theoretical formula (1). This phenomenon is called the shadow-increase, and the first to mention its size was Lahire (1707). He gave the value of 60″ at the observed times of contacts which corresponded to about 1/41 of the shadow's radius. Later mention of the shadow-increase may be found in the works of J. D. Cassini (1740), who attributed it to the influence of the Earth's atmosphere, and also in those of Lemonnier (1746), Lalande (1783), and Lambert (1782). However, a method, for the determination of this quantity was not discovered by any of these authors. Legentil (1755) even found a certain flattening of the shadow which, however, was opposed to its actual direction—a similar and as erroneous opinion as that prevailing at the same time concerning the geoid.

The first one to develop a mathematical method for the determination of the shadow-increase was Maedler (Beer and Maedler, 1834). His method was based on the length of the chord circumscribed by a certain crater within the shadow, as derived from the duration of the eclipse. Related to his work was that of Schmidt (1856) who used numerous observations of his own. The first attempt at a complete solution of this problem was made by Brosinsky (1888) who, by the

application of Maedler's method, treated twenty-three eclipses observed from 1776 to 1888. This method is advantageous, since it does not require an accurate knowledge of the moments of ingress and emergence of the craters to and from the shadow, but only of the duration of their obscuration. Neither is it necessary to know accurately the longitude of the Moon—i.e., an element that has always been the weak spot of all theories of lunar motions, while the latitude and parallax are usually known with an accuracy amply sufficient for this purpose. This method is, of course, confined to cases in which the whole course of the phenomenon has been observed.

After Brosinsky, Hartmann (1891) analyzed twenty-eight eclipses between 1802 and 1889. He did not, however, take into account the duration of the eclipse for the individual craters, but for each contact of the crater with the border of the shadow he computed its distance from the center of the shadow, and thus, simultaneously, its enlargement. This method requires, of course, a good knowledge of the moments of contact, and assumes, moreover, a reliable ephemeris of the Moon. Its advantage consists in the possibility of treating even incomplete observational series, when either the ingress or egress of some of the craters are missing.

All these works used the ecliptical coordinates current during earlier times. Today, for modern investigations of eclipses, the equatorial system in which detailed hourly ephemerids of the Moon are given, is more appropriate. It was by Kosik (1940) that our problem was treated in this system. His method, very neat from the mathematical point of view and very clear for the computer, is as follows:

Let us locate the origin of a rectangular system of coordinates at the center of the Earth, so that the axis Oz will be directed along the axis of the shadow, taken positively in the direction of the solar rays. The Oy-axis will be in the direction of the celestial north pole, and Ox lies in the equatorial plane in the direction of the lunar motion round the Earth. In this geocentric system, the coordinates of the Moon will be

$$\left.\begin{aligned} x_{\text{☾}} &= \frac{\cos\delta_{\text{☾}}\sin(\alpha_{\text{☾}}-\alpha_1)}{\sin\pi_{\text{☾}}}, \\ y_{\text{☾}} &= \frac{\sin(\delta_{\text{☾}}-\delta_1)}{\sin\pi_{\text{☾}}} + 0.008726(\alpha_{\text{☾}}-\alpha_1)\,x_{\text{☾}}\sin\delta_1, \\ z_{\text{☾}} &= \frac{\cos(\delta_{\text{☾}}-\delta_1)}{\sin\pi_{\text{☾}}} - 0.008726(\alpha_{\text{☾}}-\alpha_1)\,x_{\text{☾}}\cos\delta_1, \end{aligned}\right\} \quad (2)$$

where $\alpha_{\text{☾}}$, $\delta_{\text{☾}}$ are the apparent equatorial coordinates of the Moon, α_1, δ_1 analogous coordinates of the antisun, and $\pi_{\text{☾}}$ the lunar parallax.

Moreover, we shall introduce the selenographic system of coordinates, in

which each crater is characterized by the spherical coordinates β, λ and the distance $R_☾$ from the center of the Moon, or the rectangular coordinates

$$\left.\begin{aligned} x_0 &= R_☾ \cos\beta \sin\lambda, \\ y_0 &= R_☾ \sin\beta, \\ z_0 &= R_☾ \cos\beta \cos\lambda. \end{aligned}\right\} \quad (3)$$

These selenographic coordinates will be transformed into the geocentric system previously mentioned by means of the respective translation of origin and the rotation of the axes of coordinates according to the equations

$$\left.\begin{aligned} x &= x_☾ + a_x x_0 + b_x y_0 + c_x z_0 , \\ y &= y_☾ + a_y x_0 + b_y y_0 + c_y z_0 , \end{aligned}\right\} \quad (4)$$

where the direction cosines of the axes Ox, Oy, and Oz of the first system in the second system are given by relations

$$\left.\begin{aligned} a_x &= -\cos\lambda_\odot \cos P + \sin\lambda_\odot \sin P \sin\beta_\odot , \\ b_x &= \sin P \cos\beta_\odot , \\ c_x &= \sin\lambda_\odot \cos P - \cos\lambda_\odot \sin P \sin\beta_\odot , \\ a_y &= \cos\lambda_\odot \sin P - \sin\lambda_\odot \cos P \sin\beta_\odot , \\ b_y &= \cos P \cos\beta_\odot , \\ c_y &= -\sin\lambda_\odot \sin P - \cos\lambda_\odot \cos P \sin\beta_\odot , \end{aligned}\right\} \quad (5)$$

in which $\lambda_\odot$, $\beta_\odot$ are the selenographic coordinates of the Sun, and P the angle of position of the lunar rotation axis projected into the plane perpendicular to the axis Oz and passing through the center of the Moon. These quantities as well as α, δ, π may be found in the current ephemerides.

It is sufficient to compute all the above-mentioned quantities for hourly intervals and to interpolate them, in addition, for the observed moment of contact of the crater with the edge of the shadow. The distance of the crater from the center of the shadow will, then, be found from the equation

$$r = (x^2 + y^2)^{1/2}, \quad (6)$$

and the position angle computed from the equator

$$\log \psi = y/|\,x\,|. \quad (7)$$

Owing to the fact that the crater lies always nearer to the Earth than the center of the Moon, the following slight correction

$$\Delta r = 0.0046(z - z_{☾}) \tag{8}$$

must be introduced for r, while the angle ψ does not change appreciably. The value of $z - z_{☾}$ can be estimated.

At the end of the 19th century, when Hartmann's results (1891) concerning the enlargement of the Earth's shadow had been published, attempts were made to explain this phenomenon by the course of the illumination in the vicinity of the shadow's edge, where it very abruptly changes with the distance from the center of the shadow. This was the stimulus from which the photometric theory of lunar eclipses developed. Today, this theory has become an integral part of astronomy. In what follows, its outlines, together with a comparison with other theories, will be given.

VI

THE PHOTOMETRIC MODEL OF THE ECLIPSES

We shall consider the illumination of a certain element N of the lunar plane II (Fig. 3) by the Sun which we shall imagine as projected into plane I. A certain element M of it gives in N the illumination

$$de = kTb\, dq, \tag{9}$$

where b is the surface brightness of the Sun at the point M, dq the size of its area, T the general transmission coefficient of the rays on the trajectory between M
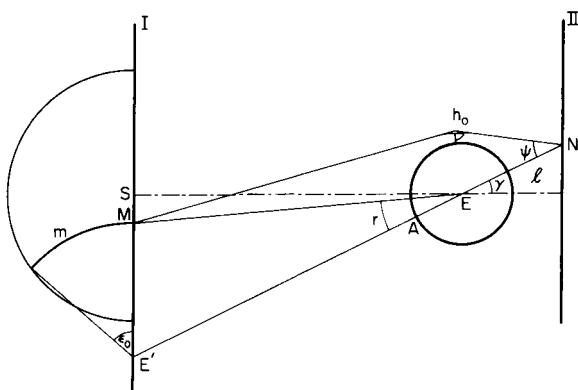
FIG. 3. Photometric diagram of lunar eclipse. (I) Solar plane, half of which is rotated into the plane of the picture, (II) lunar plane, (E) center of the Earth.

and N, and k the constant of proportion which, during one and the same eclipse, practically does not change. The total illumination at the point N will be obtained by integration

$$e = k \int Tb\, dq \tag{10}$$

over the whole solar disk.

In order to be able to carry out the integration, a so-called auxiliary shadow will be introduced. At the point N, where the illumination by the Sun is unknown, we assume a luminous point projecting the Earth's shadow—the auxiliary shadow—into the solar plane. Let the Sun be substituted by a diffusing disk, the albedo of which varies just as the brightness of the solar disk decreases towards its limb. The total light scattered by this disk towards A is, then, proportional to the source for illumination at the point N, provided that the value of the transmission coefficient T is independent of the direction of the ray, as will be proved later (Section VII).

In this way, the illumination integral (10) may be expressed more explicitly. As integration element, by which the whole solar surface will be exhausted, we shall choose a ring of the radius of $E'M$ corresponding to the geocentric angle r (Fig. 3). This ring subtends a mantle of a cone from the rays MN rotated round the axis NEE'. On a certain mantle, the transmission coefficient T is, then, constant, being only a function of the minimum altitude of the rays h_0, or of the angle r.

The light flux from the ring of the radius r and thickness dr will be (Fig. 3)

$$di = 2 \int_0^{\epsilon_0} b(R) r\, dr\, d\epsilon. \tag{11}$$

The attenuation of this flux owing to its passage through the atmosphere will be $T(r)$, so that the illumination integral will be transformed into the expression

$$e = k \int_{\gamma - R_\odot}^{\gamma + R_\odot} T(r)\, di, \tag{12}$$

where $\gamma = \sphericalangle SEE'$ is the geocentric distance of the point N from the center of the shadow.

Beyond the eclipse, $T(r) = 1$, so that the illumination is reduced to

$$E = k \int_{\gamma - R_\odot}^{\gamma + R_\odot} di. \tag{13}$$

The shadow density at the point N is, then, given by the expression

$$D = \log_{10}(E/e). \tag{14}$$

This clearcut procedure of treating the complex problem of the eclipse has not always been applied. Hepperger (1895), for instance, and after him Seeliger (1896), in integrating the illumination, proceeded from the view at the eclipse as it appears to the lunar observer. This picturesque—but from our point of view very involved procedure—will be described later (Section XII). The same path has, later on, been entered by Saussure (1931). Fessenkov (1937) integrates already in the solar plane; as integration element, however, he chooses rather inopportunely a ring concentric with the Sun, as against our ring, concentric about the center of the auxiliary shadow E'.

VII

GENERAL TRANSMISSION COEFFICIENT

For the numerical evaluation of the integral of illumination (12), the general transmission coefficient $T(r)$ and the angle r must be expressed by means of the minimum altitude of the ray h_0 above the Earth's surface. In other words, the path of the rays in the terrestrial atmosphere requires a closer examination.

For the angle r it will be easy to derive from the geometrical configuration of the rays (Fig. 3) the equation

$$r = (\pi_{\odot} + \pi_{☾})[1 + (h_0'/a)] - \omega, \tag{15}$$

where $\pi_{\odot}$, $\pi_{☾}$ are the parallaxes of the Sun and Moon, h_0' the elevation of the rays' asymptote, and ω its total deviation due to the refraction in the atmosphere of the Earth.

In an ideal atmosphere (Rayleigh's atmosphere), the general transmission coefficient $T(r)$ consists of two components. The first component is the extinction

$$10^{-AM}, \tag{16}$$

where M is the air-mass or equivalent path of the ray, and A the extinction-coefficient for the given wavelength, so as it results from the theory for the molecular scattering of light (Section XVIII).

The second component is the attenuation of light by refraction which causes the normal divergence of the rays to be altered. A thin pencil of rays (Fig. 4) contained between the cones of the angles of vertex τ and $\tau + d\tau$ would—if there were no refraction—subtend a ring of the area of dS in the lunar plane II. The pencil deflected by refraction subtends, however, a ring of dS'. Since the light flux is equal in both cases, in the lunar plane an attenuation of the illumina-
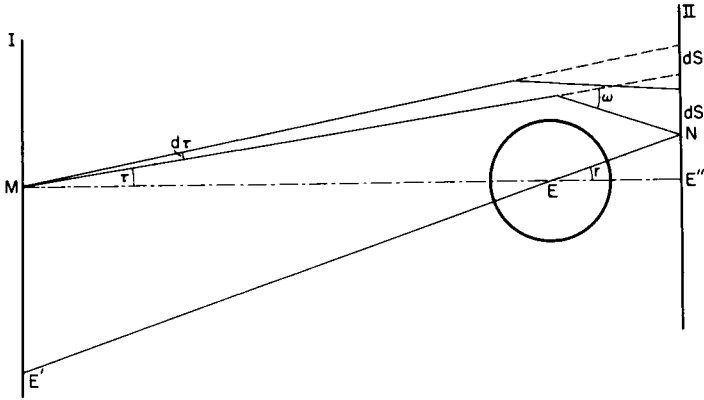
FIG. 4. Attenuation of light by refraction.

tion in the ratio $dS : dS'$ is produced. From the geometry of the rays the following expression for this ratio may be derived

$$\begin{aligned}\phi = \frac{dS'}{dS} &= \left[1 - \frac{\omega}{\pi_\odot + \pi_☾}\left(1 - \frac{h_0'}{a}\right)\right]\left[1 - a\frac{d\omega}{dh_0'}\frac{1}{\pi_☾ + \pi_\odot}\right] \\ &= \frac{r}{(\pi_☾ + \pi_\odot)\left(1 + \dfrac{h_0'}{a}\right)}\left(1 - a\frac{d\omega}{dh_0'}\frac{1}{\pi_☾ + \pi_\odot}\right).\end{aligned} \tag{17}$$

Thus, the general transmission coefficient in Rayleigh's atmosphere will become

$$T(r) = 10^{-AM\Phi^{-1}}, \tag{18}$$

which evidently does not depend on the rays' direction, as assumed in Section VI.

The expression for the attenuation of refractions consists of two parts: namely,

$$1 - \frac{\omega}{\pi_☾ + \pi_\odot}\left(1 - \frac{h_0'}{a}\right) \tag{19}$$

and

$$1 - a\frac{d\omega}{dh_0'}\frac{1}{\pi_\odot + \pi_☾}, \tag{20}$$

which act in the opposite sense.

The first part represents the increase of illumination, the second its attenuation. These circumstances may also be presented geometrically. Let a thin cylindrical pencil of rays (Fig. 5) penetrate into the terrestrial atmosphere. If it were refracted uniformly and independently of the height of the ray, the pencil—after having
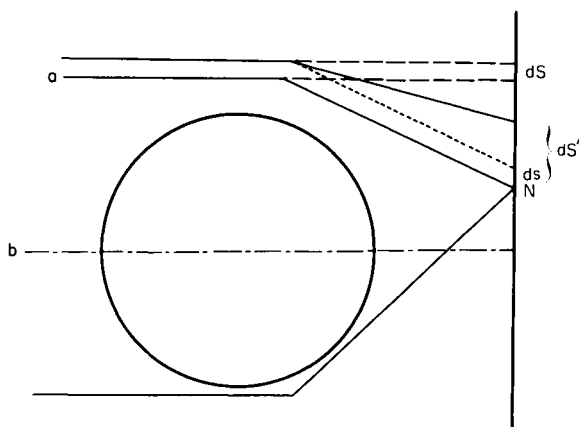
FIG. 5. Two components of light-attenuation by refraction. Double illumination at point N by the higher ray a as well as the lower ray b.

touched the lunar plane—would produce a ring ds, and an increase of illumination would ensue, which is the effect of the above-mentioned term (19). Actually, the refraction does change with the altitude and, in the plane II, a ring dS' is produced which corresponds to the attenuation of the illumination, as expressed by the second term (20).

For various rays (i.e., for various h_0's), the value of both terms changes, and so will the resulting effect. For very high rays (i.e., for small refractions), we may write (29)

$$d\omega/dh_0' = -\beta\omega, \tag{21}$$

and the effect of refraction is then reduced to the simple expression

$$\Phi = 1 + \frac{a\beta\omega}{\pi_{☾} + \pi_{\odot}}. \tag{22}$$

This expression was already known to Pannekoek (1903) and Fabry (1929) who studied occultations of stars by planets. Elsewhere (Dollfus, 1962) we have already seen the application of a similar phenomenon for a research of the lunar atmosphere.

Now, if we descend deeper into the atmosphere, the second term retains its predominance, and the resulting effect is a continuously growing attenuation with a maximum for rays around $h_0 = 13$ km (Table III). Lower down, the attenuation decreases; at an altitude of 2 km it equals unity, and for rays for which condition

$$\omega = (\pi_{☾} + \pi_{\odot})[1 + (h_0'/a)] \tag{23}$$

is fulfilled, in other words, where $r = 0$ (i.e., in the center of the auxiliary shadow) the focusing of the rays takes place. This fact was known already to Laplace (1860). In this case, the altitude of the rays is a little lower than 2 km.

For still lower rays, r as well as Φ becomes negative. This is, however, only of geometrical significance. Into point N of the auxiliary shadow (Fig. 5), the rays now arrive along path b—i.e., from the antipodes of the point of the higher rays a which previously were incident on the point N. The inner parts of the auxiliary shadow are thus illuminated twice as far as to the distance $r = -17'$ (for the mean lunar parallax 57′). The effect of the second illumination is, of course, practically negligible; since atmospheric layers below 2 km of altitude are very little transparent and, moreover, their transparence may be impaired by clouds.

VIII

COMPUTATION OF THE REFRACTION AND OF THE AIR-MASS

For practical applications of the foregoing equations, the knowledge of the refraction ω of the air mass M is necessary for horizontal rays as functions of their minimum altitude h_0. A similar problem has already been dealt with by classical astronomy in numerous variants of the refraction theory. There is, however, one basic difference. Classical astronomy computes the refraction and, eventually, the air mass, as a function of the zenith distance for the vertex of the ray in a low altitude above sea level; while in our case, the argument is precisely the elevation of the vertex of the horizontal rays extending to considerable altitudes. The classical refraction theories mostly fail, or attain only a limited accuracy at the horizon—i.e., just where a knowledge of the refraction is required for our purposes.

The principal cause of this deficiency of classical theories rests in the approximations of the function which represents the variation of the air-density ρ with the altitude h. This generally complicated function was represented by such analytical approximations as to enable us to evaluate the refraction integral by a development into a series. In classical astronomy, this method was adequate as, according to the well-known Oriani–Laplace theorem, the refraction is practically independent of the structure of the atmosphere to considerable zenith distances.

In the problem of lunar eclipses, if the results are to be compared with those of observations, we must approximate the reality more closely. Therefore, we shall use only the numerical form of function $\rho = f(h)$ as given, for instance, by the average results of aerological or rocket soundings of the atmosphere. With this air density the refraction index is then associated through the well-known relation

$$\mu = 1 + c\rho, \tag{24}$$

where $c = 293 \times 10^{-6}$ for the center of the visible spectrum.

If we designate the angle subtended in a general point by its radiusvector with the ray by i we obtain, according to the invariant theorem,

$$\mu(1 + h) \sin i = \mu_0(1 + h_0) \sin z. \tag{25}$$

The total deviation of the ray, that is, the angle ω is given by the integral (Link, 1933)

$$\omega = 2 \int_0^1 K \, dZ, \tag{26}$$

where the independent variable is given by the equation

$$Z = \cos i \tag{27}$$

and

$$K = c\rho\beta/(1 - c\rho\beta), \tag{28}$$

where

$$\beta = \frac{d\rho/dh}{\rho} \tag{29}$$

denotes the gradient of density with the altitude.

Similarly, for the air mass M we obtain (Link, 1933)

$$M = 2 \int_0^1 (K/c\beta) \, dZ. \tag{30}$$

In this form, ω and M may be computed by numerical integration. Let us depart from the numerical form of function $\rho = f(h)$ given, for instance, in the form of a table of values for the equidistant values h (for instance, by 1 km). From this function we derive quantity β. Then, by means of Eqs. (27)–(29), the functions behind the sign of integration of the integrals (26) and (30) may be computed, and, finally, also the values of these integrals. Their form is generally given by the curve in Fig. 6, from which it is evident that the greatest influence
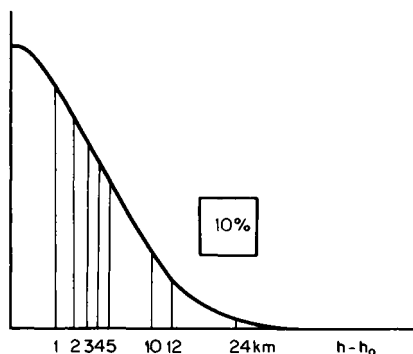
Fig. 6. General form of the refraction integral.

on the magnitude of the integral is exerted by the structure of the atmosphere in the vicinity of the vertex of the ray's path (i.e., at the altitude h_0 and some 10 km above). Higher altitudes enter the integral only with relatively small weight. Should we stop the integration at an altitude of $h_0 + 12$ km, this would entail an error of only about 5% of the total value of the integral. If, on the other hand, the values from the altitude $h_0 + 12$ km on are subject to an error up to 20%, the error of the whole integral will be less than 1%. As will be shown in the following section, our knowledge of the atmosphere is sufficient for a computation of ω and M within the whole range of the shadow and penumbra. Here, the structure of higher atmospheric strata appears only as correction term.

IX

BRIGHTNESS OF THE SOLAR ELEMENTARY RING

The first photometric theories of lunar eclipses neglected the limb darkening of the solar disk. The brightness of this latter disk is, however, known to decrease from the center towards its limb about twice as fast in the violet as in the red region of the spectrum. If we neglect it, the resulting densities of the shadow will be somewhat lower—particularly at the edge of the shadow, where the effective part of the solar disk is reduced to a narrow crescent.

Later, beginning with Seeliger (1896), this important factor has been taken into account by the photometric theories. In our theory, there appears the light intensity of a ring of the width of dr subtended on the Sun by the radius r from the center, at a fractional distance γ from the center of the Sun. In accordance with Eq. (11), the general expression for it becomes

$$di = 2 \int_0^{\epsilon_0} b(R) r \, dr \, d\epsilon, \tag{11}$$

where the function $b(R)$ defines the distribution of brightness over the solar apparent disk. If we approximate it by the usual expression,

$$b(R) = 1 - \kappa + (\kappa/R_\odot)(R_\odot^2 - R^2)^{1/2}, \qquad \kappa < 1, \tag{31}$$

where κ is a constant dependent on the wavelength of the light. This relation will prove true up to 97% $R_\odot$. Closer to the limb, the actual brightness decreases more rapidly as was shown for instance by Heyden (1954) in his measurements during the eclipse of the Sun.

In various cases mentioned below we shall need the luminosity of the following elementary rings of the Sun:

TABLE II

LUMINOSITY OF SOLAR ELEMENTS [CASES (b) AND (c)][a]

	$\gamma = 0'$		$\gamma = 5'$		$\gamma = 10'$		$\gamma = 20'$		$\gamma = 30'$		$\gamma = 36'$		$\gamma = 40'$		$\gamma = 45'$		$\gamma = 50'$		$\gamma = 58'$		$\gamma = 68'$	
$r - \gamma^b$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$	$P/2$	$Q/2$
15	47.1	30.7	11.3	8.2	8.8	6.4	7.9	5.9	6.8	5.0	6.6	4.8	6.5	4.7	6.4	4.6	6.3	4.6	6.2	4.5	6.1	4.5
14	44.0	22.7	15.5	9.6	12.1	7.5	10.1	6.3	9.4	5.8	9.1	5.7	9.0	5.6	8.9	5.5	8.8	5.4	8.6	5.3	8.5	5.3
13	40.8	17.0	18.5	10.1	14.4	7.8	12.0	6.5	11.2	6.1	10.9	5.9	10.8	6.0	10.6	5.7	10.5	5.6	10.3	5.5	10.2	5.4
12	37.7	12.8	20.8	10.2	16.0	7.7	13.5	6.5	12.6	6.0	12.2	5.9	12.1	5.8	11.9	5.7	11.8	5.7	11.6	5.6	11.5	5.5
11	34.6	9.5	22.6	10.0	17.3	7.5	14.6	6.3	13.7	5.9	13.3	5.7	13.2	5.7	13.0	5.6	12.8	5.4	12.7	5.3	12.5	5.3
10	31.4	6.9	24.2	9.9	18.3	7.2	15.5	6.0	14.5	5.6	14.2	5.5	14.0	5.4	13.8	5.3	13.7	5.3	13.5	5.2	13.4	5.2
9	28.3	4.9	25.5	9.7	19.0	6.8	16.1	5.7	15.2	5.3	14.9	5.2	14.7	5.2	14.5	5.1	14.4	5.0	14.2	5.0	14.1	4.9
8	25.1	3.4	26.9	9.9	19.5	6.4	16.6	5.3	15.7	5.0	15.4	4.9	15.2	4.9	15.1	4.8	14.9	4.8	14.8	4.7	14.6	4.7
7	22.0	2.2	28.6	10.6	19.8	6.0	17.0	4.9	16.1	4.7	15.8	4.7	14.7	4.6	15.5	4.6	15.4	4.5	15.2	4.5	15.1	4.4
6	18.8	1.4	34.6	14.2	20.0	5.6	17.2	4.8	16.4	4.3	16.1	4.4	16.0	4.4	15.8	4.3	15.7	4.3	15.6	4.2	15.4	4.2
5	15.7	0.8	31.4	9.9	20.1	5.4	17.3	4.5	16.6	4.2	16.3	4.2	16.2	4.1	16.1	4.1	15.9	4.0	15.8	4.0	15.7	4.0
4	12.6	0.4	28.3	7.2	20.0	5.2	17.3	4.2	16.7	4.0	16.5	4.0	16.3	3.9	16.2	3.9	16.1	3.8	16.0	3.8	15.9	3.8
3	9.4	0.2	25.1	5.1	19.8	4.9	17.3	4.0	16.6	3.8	16.5	3.8	16.4	3.8	16.3	3.8	16.3	3.8	16.2	3.7	16.0	3.7
2	6.3	0.0	22.0	3.6	19.5	4.8	17.0	3.8	16.6	3.6	16.4	3.6	16.4	3.6	16.3	3.6	16.2	3.6	16.1	3.6	16.1	3.6

+1	3.1	0.0	18.8	2.5	19.0	4.6	16.8	3.7	16.4	3.6	16.3	3.6	16.3	3.5	16.2	3.5	16.1	3.5	16.1	3.5	16.1	3.5
0	0.0	0.0	17.7	1.7	18.5	4.6	16.5	3.6	16.2	3.5	16.1	3.5	16.1	3.5	16.1	3.5	16.0	3.4	16.0	3.4	16.0	3.4
−1			12.6	1.1	18.0	4.6	15.8	3.6	15.9	3.5	15.9	3.5	15.9	3.4	15.9	3.5	15.8	3.4	15.8	3.4	15.9	3.4
2			9.4	0.7	17.5	5.0	15.5	3.5	15.5	3.5	15.6	3.5	15.6	3.5	15.6	3.5	15.6	3.4	15.6	3.4	15.6	3.5
3			6.3	0.4	17.1	5.5	15.0	3.5	15.1	3.5	15.2	3.5	15.2	3.5	15.3	3.5	15.2	3.5	15.3	3.5	15.4	3.5
4			3.1	0.2	16.8	5.2	14.3	3.5	14.6	3.5	14.7	3.6	14.8	3.6	14.9	3.6	14.8	3.6	14.9	3.6	15.0	3.6
5			0.0	0.0	15.7	5.0	13.6	3.5	14.0	3.7	14.2	3.6	14.3	3.7	14.4	3.7	14.4	3.7	14.5	3.7	14.6	3.7
6					12.6	3.5	12.9	3.6	13.4	3.7	13.7	3.7	13.8	3.8	13.9	3.8	13.9	3.8	14.0	3.8	14.2	3.9
7					9.4	2.5	12.0	3.6	12.7	3.8	13.0	3.8	13.1	3.9	13.3	3.9	13.3	3.9	13.5	4.0	13.6	4.0
8					6.3	1.5	11.1	3.6	12.0	3.9	12.3	3.9	12.5	4.0	12.6	4.0	12.7	4.1	12.9	4.1	13.0	4.2
9					3.1	0.7	10.2	3.6	11.2	3.9	11.5	4.0	11.7	4.1	11.9	4.2	12.0	4.2	12.1	4.4	12.3	4.3
10					0.0	0.0	9.1	3.6	10.3	4.0	10.7	4.2	10.9	4.2	11.0	4.3	11.2	4.3	11.4	4.4	11.6	4.5
11							7.8	3.4	9.3	4.0	9.7	4.2	9.9	4.3	10.1	4.4	10.3	4.3	10.5	4.4	10.7	4.5
12							6.9	3.3	8.3	4.2	8.7	4.2	8.9	4.3	9.1	4.3	9.2	4.4	9.4	4.5	9.6	4.6
13							5.6	3.0	7.1	3.8	7.5	4.1	7.7	4.2	7.9	4.3	8.0	4.3	8.2	4.4	8.4	4.5
14							4.3	2.7	5.7	3.5	6.0	3.7	6.3	3.9	6.4	4.0	6.6	4.1	6.7	4.2	6.9	4.3
15							2.8	2.0	3.9	2.9	4.3	3.1	4.4	3.2	4.5	3.3	4.6	3.4	4.8	3.5	4.9	3.6

[a] From Link, 1963.
[b] For $r - \gamma = \pm 16'$, we have $P = 0, Q = 0$ except for $\gamma = 0'$ where $P = Q = 100.6$.
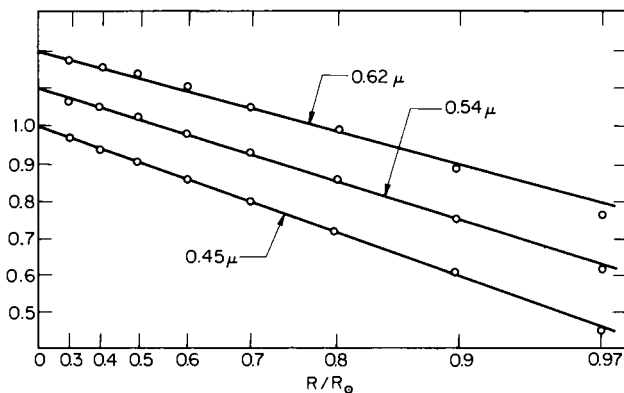
FIG. 7. Changes of brightness on the solar disk.

(a) An elementary ring described from the center of the sun by the radius R. Its luminosity will be

$$di = 2\pi R b(R)\, dR. \tag{32}$$

(b) An elementary ring described by the radius r from the center E' lying at a distance γ from the center of the sun for $\gamma \geqslant R_\odot$ (Fig. 3),

$$di = 2\int_0^{\epsilon_0} (1 - \kappa) r\, dr\, d\epsilon + 2\int_0^{\epsilon_0} (\kappa r/2R_\odot)(2r\gamma)^{1/2}(\cos\epsilon - \cos\epsilon_0)^{1/2}\, dr\, d\epsilon$$

or

$$di = 2[(1 - \kappa)\,\epsilon_0 r + (\kappa r/R_\odot)(r\gamma)^{1/2} I(m)]\, dr \tag{33}$$

where

$$I(m) = \frac{\pi m\sqrt{2}}{8}\left(1 + \frac{m}{32} + \frac{3m^2}{1024} + \cdots\right),$$
$$m = 4\sin^2\frac{\epsilon_0}{2} = \frac{(R_\odot + \gamma - r)(R_\odot + r - \gamma)}{r\gamma}. \tag{34}$$

(c) An elementary ring of analogous properties, for $\gamma < R_\odot$ and $r < R_\odot - \gamma$, is

$$di = 2\pi r(1 - \kappa)\, dr + \frac{2\kappa}{R_\odot}[R_\odot^2 - (\gamma - r)^2]^{1/2}$$
$$\times \int_0^{\pi}\left(1 - \frac{4r\gamma}{R_\odot^2 - (\gamma - r)^2}\sin^2\frac{\epsilon}{2}\right)^{1/2} d\epsilon\, dr \tag{35}$$

where the integral on the right-hand side is an elliptic integral of second kind

$$\int_0^{\pi} \left(1 - \frac{4r\gamma}{R_{\odot}^2 - (\gamma - r)^2} \sin^2 \frac{\epsilon}{2}\right)^{1/2} d\epsilon = 2E\left(\vartheta, \frac{\pi}{2}\right) \qquad (35')$$

$$\sin^2 \vartheta = 4r\gamma/[R_{\odot}^2 - (\gamma - r)^2].$$

X

STRUCTURE OF THE AUXILIARY SHADOW

The computation of the structure of the auxiliary shadow presumes the knowledge of function $\rho = f(h)$. This computation was carried out (Link, 1933, 1958) for the atmosphere and lower stratosphere to an altitude of 26 km form Humphreys' data (1929), and higher up to 120 km according to data from

TABLE III

AUXILIARY SHADOW

h (km)	ω ′	M[a]	r ′	d_1	d_2	d_3	d_4	D_2	D_3	D_4
2	55.7	60.6	1.4	1.30	2.79	0.32	0.12	4.09	4.41	4.53
5	40.5	41.9	16.6	2.18	1.92	0.33	0.12	4.11	4.44	4.56
10	25.2	21.4	32.0	2.24	0.98	0.33	0.12	3.22	3.55	3.67
15	12.2	9.8	45.0	2.24	0.45	0.33	0.12	2.69	3.02	3.14
20	5.6	4.4	5.16	1.95	0.20	0.32	0.12	2.15	2.47	2.59
25	2.6	2.0	54.7	1.64	0.09	0.29	0.12	1.73	2.02	2.14
30	1.4	0.9	56.2	1.31	0.04	0.17	0.12	1.35	1.52	1.64
35	0.5	0.5	56.8	1.00	0.02	0.07	0.11	1.02	1.10	1.21
40	0.2	0.2	57.3	0.70	0.01	0.01	0.10	0.70	0.71	0.81
50	0.1	0.1	57.5	0.26	0.00	0.00	0.10	0.26	0.26	0.36
60	0.0	0.0	57.7	0.09	0.00	0.00	0.08	0.09	0.09	0.17
70	0.0	0.0	57.8	0.02	0.00	0.00	0.08	0.02	0.02	0.10
80	0.0	0.0	57.9	0.00	0.00	0.00	0.06	0.00	0.00	0.06
90	0.0	0.0	57.9	0.00	0.00	0.00	0.04	0.00	0.00	0.04
100	0.0	0.0	58.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00

[a] In 8 km normal atmosphere units.

Computed for $\pi_☾ = 57'$ and $\lambda = 5400$ Å.

d_1 the attenuation by refraction,
d_2 the extinction in the Rayleigh's atmosphere,
d_3 the absorption by ozone,
d_4 the absorption by high absorbing layer (Section XVIII).

$D_2 = d_1 + d_2$, $D_3 = D_2 + d_3$, $D_4 = D_3 + d_4$.

Rocket Panel (1952). These data made it possible satisfactorily to compute the structure of the auxiliary shadow for mean latitudes and for the mean season of the year as far as to its limit 58′. An excerpt from the results for the mean parallax $\pi_{☾} = 57'$ is presented in Table III.

This table is instructive from several points of view. If we observe the way in which the component d_1 arising from the attenuation by refraction and the component of extinction d_2 are represented in the total density of the auxiliary shadow, we see that in the borderline regions, it is attenuation by refraction that plays the predominant part. This is independent of the frequency of light, so that the borderline regions, too, will be neutral in color. In the internal regions, the component of extinction—strongly selective—begins to prevail; and this accounts for the reddish coloring of the auxiliary shadow. At the very center of the shadow, there is an apparent decrease of the component d_1 due to focusing of the rays, also reflected in the total density.

The density of the actual shadow is computed by integration (Eq. (12)) from the extended region ($2R_{\odot}$) of the auxiliary shadow. The above-mentioned characteristics of the auxiliary shadow are also transferred into the actual shadow, even if with a lesser contrast—a fact which is fully confirmed by the observations.

XI

NORMAL DENSITIES OF THE SHADOW

The normal densities of the shadow have been computed (Link, 1933, 1948a) for Rayleigh's atmosphere in three spectral regions—the red, green, and blue—

TABLE IV

Color		c	A	κ
Red	6200 Å	292.0×10^{-6}	0.0263	0.54
Green	5400 Å	293.2×10^{-6}	0.0460	0.62
Blue	4600 Å	295.5×10^{-6}	0.0889	0.73

under application of the following constants, for that of three values 54′, 57′, 61′ of the lunar parallax. The results are presented in Table V.

The numerical results indicate one useful characteristic of the shadow: namely, that the structure of its limb is practically independent of the Moon's parallax when, instead of the angular distance from the center of the shadow γ, we introduce as argument the distance from the edge of the shadow, defined by relation

$$\gamma'' = \pi_{☾} - R_{\odot} - \gamma. \tag{36}$$

TABLE V

NORMAL DENSITIES OF THE SHADOW[a]

		Densities						Mean air-masses					
	$\pi_☾$	$\gamma = 0'$	5′	10′	20′	30′	35′	0′	5′	10′	20′	30′	35′
	54′	3.10	3.09	3.06	2.97	2.76	2.58	49.5	48.4	43.5	27.8	17.5	11.8
	55′	3.14	3.13	3.10	3.01	2.79	2.62	50.3	49.5	44.4	28.7	18.9	13.1
	56′	3.18	3.16	3.14	3.05	2.81	2.66	51.2	50.6	45.3	29.8	20.1	14.3
$\lambda = 0.62\,\mu$	57′	3.21	3.20	3.17	3.09	2.83	2.70	52.2	51.7	46.3	31.1	21.2	15.5
	58′	3.25	3.24	3.21	3.12	2.85	2.73	53.4	52.6	47.4	32.3	22.1	16.5
	59′	3.29	3.27	3.24	3.16	2.87	2.76	54.5	53.4	48.7	33.7	22.8	17.5
	60′	3.32	3.31	3.27	3.18	2.90	2.78	55.5	54.2	50.0	35.2	23.5	18.4
	61′	3.36	3.35	3.31	3.20	2.92	2.80	56.7	54.9	51.8	36.8	24.1	19.3
	54′	4.10	4.04	3.91	3.49	3.09	2.79	44.9	43.3	37.8	24.7	15.3	9.9
	55′	4.17	4.11	3.97	3.53	3.13	2.86	45.8	44.0	38.6	25.5	16.2	11.0
	56′	4.24	4.18	4.03	3.58	3.17	2.91	46.9	44.9	39.6	26.5	17.1	12.1
$\lambda = 0.54\,\mu$	57′	4.30	4.24	4.09	3.63	3.21	2.97	48.0	46.1	40.8	27.7	18.1	13.2
	58′	4.36	4.31	4.15	3.68	3.25	3.02	49.2	47.6	42.2	29.0	19.0	14.2
	59′	4.43	4.37	4.22	3.74	3.29	3.07	50.5	49.2	43.8	30.2	19.9	15.2
	60′	4.49	4.43	4.28	3.80	3.33	3.12	52.0	50.8	45.5	31.5	20.8	16.1
	61′	4.55	4.49	4.35	3.87	3.37	3.16	53.3	52.7	47.3	32.9	21.7	17.0
	54′	5.86	5.69	5.30	4.46	3.68	3.18	42.0	37.6	32.1	20.9	12.5	7.7
	55′	5.98	5.77	5.40	4.55	3.76	3.28	43.3	39.2	33.0	21.7	13.2	8.5
	56′	6.10	5.87	5.50	4.63	3.85	3.38	44.8	40.8	34.1	22.7	14.1	9.5
$\lambda = 0.46\,\mu$	57′	6.23	5.98	5.60	4.72	3.93	3.48	46.3	42.4	35.2	23.7	15.1	10.5
	58′	6.36	6.11	5.72	4.82	4.01	3.57	47.8	43.8	36.6	24.8	16.2	11.4
	59′	6.49	6.25	5.84	4.93	4.09	3.66	49.3	45.4	38.1	25.9	17.4	12.4
	60′	6.63	6.40	5.97	5.05	4.17	3.75	50.8	47.0	39.5	27.1	18.8	13.3
	61′	6.76	6.56	6.11	5.16	4.24	3.84	52.3	48.4	41.1	28.3	20.3	14.1

[a] From Link (1933), p. 77.

This property of the shadow is associated with the fact that, in the auxiliary shadow, the solar radius is rather small when compared with the curvature of the isophotes—particularly in the outer parts. Nearer towards the center, the influence of the parallax becomes greater, so that with a decreasing parallax (increasing distance) the brightness of the eclipse increases which—besides—also follows from the geometrical conditions (Table IV).

For further discussion it will be useful to investigate the form of the integral of illumination as defined by Eq. (12). This form is presented on Fig. 8 for several
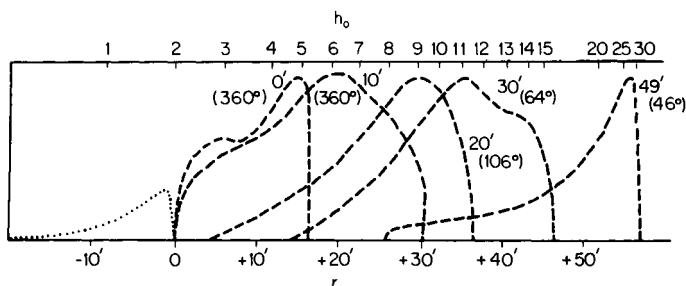
FIG. 8. Forms of the illumination integrals for $\pi_{☾} = 57'$. The maxima are normalized to 100. The upper scale indicates the altitudes of the ray h_0, the lower scale the distances r from the center of the auxiliary shadow. The figures at the curves indicate the distance γ from the center of the shadow in minutes of arc and the angular amplitude of the terminator of the shadow in degrees.

points of the shadow. These curves clearly indicate the effect of various atmospherical layers, or of clouds, on the brightness of the shadow in its individual parts. Thus, for instance, at the edge of the shadow the cloudiness of the troposphere (up to 11 km) has a negligible effect. From the distance of 30′ on (that is 11′ from the edge of the shadow), lower layers, too, begin to assert themselves. Local cloudiness has, however, only a slight effect, since the angular amplitude of the effective part of the terminator increases towards the center of the shadow. At a distance of 30′, this amplitude exceeds the quadrant (10,000 km), so that only the average cloudiness along this long track is relevant in this connection. Thus, it may be expected that the outer parts of the shadow will reflect the structure of rather limited parts of the high atmosphere, while the intermediate and central parts of the shadow will be influenced by the global transparency of the low atmospheric layers.

The form of the curves in the outer regions of the shadow is also of importance. It is apparent from them that even a comparatively thin pencil of the highest rays exerts a decisive effect on the illumination in these parts of the shadow. Such a pencil has, then, the character of tangential sounding rays, which is useful for the understanding of the structure of the upper atmosphere.

TABLE VI

NORMAL DENSITIES OF THE SHADOW AND THE MEAN AIR-MASS[a]

	$\pi_{☾}$	$\gamma'' = 0'$	1′	2′	3′	4′	5′	6′	7′	8′	9′	10′
	54′	2.37	2.46	2.53	2.58	2.63	2.67	2.70	2.73	2.76	2.79	2.81
		7.3	9.2	10.7	11.8	13.0	14.0	15.2	16.3	17.5	18.5	19.5
$\lambda = 0.62\ \mu$	57′	2.35	2.46	2.52	2.58	2.63	2.67	2.70	2.73	2.75	2.78	2.81
		7.3	8.8	10.2	11.7	13.0	14.5	15.5	16.7	17.8	19.0	20.1
	61′	2.33	2.42	2.49	2.55	2.60	2.64	2.67	2.71	2.75	2.77	2.80
		7.3	8.3	9.5	10.6	11.7	13.1	14.5	15.7	17.0	18.4	19.3
	54′	2.50	2.62	2.72	2.79	2.86	2.93	2.99	3.04	3.08	3.13	3.17
		6.0	7.4	8.8	9.9	11.1	12.1	13.2	14.3	15.3	16.4	17.4
$\lambda = 0.54\ \mu$	57′	2.48	2.61	2.70	2.78	2.85	2.92	2.97	3.02	3.07	3.13	3.17
		5.7	7.1	8.4	9.7	11.0	12.1	13.2	14.3	15.3	16.3	17.2
	61′	2.46	2.58	2.68	2.76	2.83	2.90	2.95	3.01	3.07	3.12	3.16
		2.9	7.2	8.4	9.7	11.0	12.1	13.2	14.3	15.2	16.2	17.0
	54′	2.73	2.91	3.06	3.18	3.29	3.39	3.49	3.59	3.68	3.77	3.85
		4.1	5.4	6.6	7.7	8.6	9.5	10.5	11.4	12.5	13.4	14.4
$\lambda = 0.46\ \mu$	57′	2.71	2.89	3.04	3.16	3.28	3.38	3.48	3.58	3.67	3.76	3.84
		4.1	5.4	6.6	7.7	8.6	9.5	10.5	11.4	12.4	13.3	14.2
	61′	2.69	2.86	3.01	3.16	3.26	3.37	3.46	3.56	3.65	3.75	3.84
		4.3	5.4	6.4	7.5	8.5	9.5	10.5	11.4	12.4	13.3	14.1

[a] From Link (1933).

In the preceding tables, moreover, the so-called mean air-mass M_0 is presented. The integral of illumination [(Eq. (12)] may be written

$$e = k \int 10^{-AM} \frac{dS}{dS'} di. \tag{12'}$$

The slight change of the extinction-coefficient of the atmosphere A causes a change of the illumination

$$\frac{de}{dA} = -2.3026k \int 10^{-AM} \frac{dS}{dS'} di\, dM \tag{37}$$

or a change of the shadow-density

$$\frac{dD}{dA} = \frac{\int 10^{-AM} (dS/dS')\, di\, dM}{\int 10^{-AM}(dS/dS')\, di} = M_0 . \tag{38}$$

This expression has the dimension of the air-mass and is called the mean air-mass. It is very useful in the analysis of the differences observation–computation, as will be shown later.

For illustration, Table VII presents a comparison with the results of other theories, all computed for the green light and for Rayleigh's atmosphere at $\pi_☾ = 57'$.

TABLE VII

COMPARISON OF DIFFERENT THEORIES

	Hepperger (1895)	Seeliger (1896)	Saussure (1931)	Fessenkov (1937)	Link (1933)
0	5.33	5.79			4.30
20	4.35				3.63
30	3.58				3.21
35	3.15			3.13	2.97
40	2.60		2.70	2.82	2.61
41	2.42	2.82	2.58	2.61	2.48

XII

THE ECLIPSE ON THE MOON

To a lunar observer, the eclipse of the Moon is, in fact, an eclipse of the Sun by the Earth. In contrast with the terrestrial eclipses of the Sun, the occulting body —the Earth—is surrounded by an atmospheric layer that substantially modifies the entire geometry of the phenomenon.

At a certain point N of the shadow on the Moon, at an angular distance γ from the center of the shadow, we see the center of the Sun projected into the distance

$$\gamma' = \frac{\pi_{☾}}{\pi_{☾} + \pi_{\odot}} \gamma \tag{39}$$

from the center of the Earth. The angle γ' is consequently slightly, about 0.25%, smaller than γ. The apparent solar radius, observed from the Moon, is also reduced in the same ratio; and the radius of the Earth will be visible under the angle

$$\psi_0 = \pi_{☾}(1 + c\rho^*), \tag{40}$$

where ρ^* is the air-density on the Earth's surface at the level of the cloud layer which restricts the passage of the solar ray.

A thin pencil of rays emanating from point N under the angle ψ subtends at the Sun an elementary ring AB visible from the Earth under the angle (Figs. 3 and 9)

$$r = (\pi_{☾} + \pi_{\odot})[1 + (h_0'/a)] - \omega; \tag{15}$$
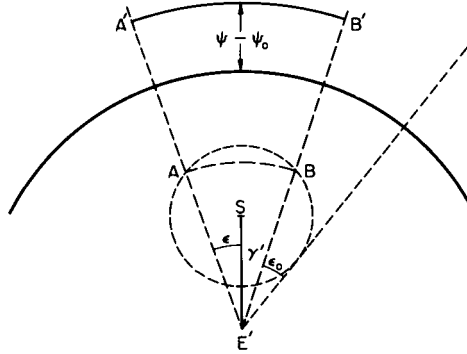
FIG. 9. Refraction image of the Sun. S is the center of the Sun, the geometrical position of which is dashed.

from the Moon, it will be visible as $A'B'$ under the angle

$$\psi = (a + h_0')/l = \pi_{☾}[1 + (h_0'/a)]. \tag{41}$$

The central angle ϵ_0 does not change. Equations (40) and (41), in fact, define the optical imaging of the solar disk by the terrestrial atmosphere. The refraction image of the whole Sun will be obtained by a superposition of a great number of the elementary images $A'B'$.

The course of the whole eclipse on the Moon in a few characteristic phases is given on Fig. 10, and may be described as follows: From the first contact between the Sun and the Earth, which corresponds to the position of the observer
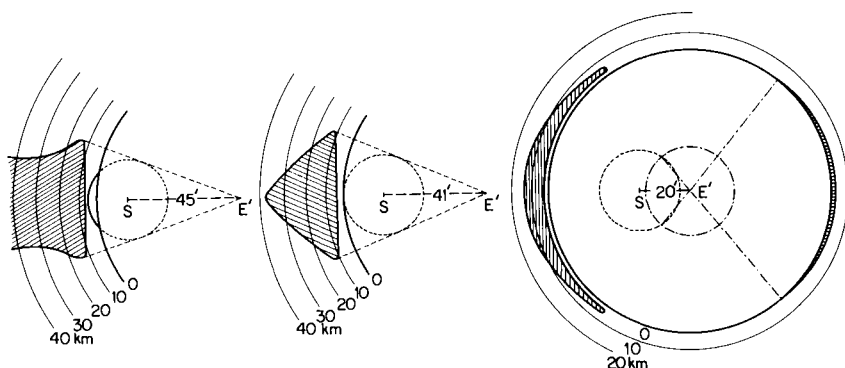
FIG. 10. Example of changes of the refraction image of the Sun. The distances from the limb of the Earth are amplified a hundred times.

N at the edge of the penumbra, the solar disk becomes gradually occulted by the opaque Earth. At the same time, the refraction image of the Sun begins to appear on both cusps of the solar crescent as well as at places symmetrically conjugate about the center E' on the opposite side of the terrestrial limb. This second refraction image originates from the lower pencil of rays incident into point N, as has been explained in more detail in Section VII.

Following the second (i.e., the internal) contact of the eclipse, on the edge and within the shadow, only both refraction images of the Sun remain visible, and continuously creep along the limb of the Earth until they blend together into a closed ring. These refraction images are, of course, only very narrow—at most only a fraction of 1′ of arc. Their brightness rapidly decreases towards the internal edge (i.e., towards the Earth) and their red coloring simultaneously intensifies. The first image on the internal side and the whole of the second image may be locally interrupted by clouds or mountain ranges. The structure of the Earth's atmosphere, varying with the geographical latitude, generally produces an increase in width of the refraction image in the direction towards the poles (as the refraction there is greater), and the altitude of the cloudy layer, restricting the internal edge of the image, is lower.

First photometrical theories were based on these "imaginary" observations from the Moon (Hepperger, 1895; Seelinger, 1896). In terms of the modern nomenclature they may be summed up by saying that the illumination at the point N of the lunar plane by the elementary ring dm' is given by the product of the average surface brightness b of the ring and the solid angle under which it appears from the point N—i.e.,

$$de = kb\, 10^{-AM}\, dm' \tag{9'}$$

where dm' is the solid angle of the ring.

The total illumination is then given by the integral

$$e = k \int b\, 10^{-AM}\, dm', \tag{10'}$$

in which the whole refraction image of the sun will be involved; with Hepperger setting $b = 1$ and Seeliger taking account of the variable brightness on the solar disk.

These theories appear to have disregarded the attenuation by refraction which is, however, implicitly comprised within the solid angle dm'. The latter is reduced for attenuation by refraction, unlike the solid angle of the ring dm, as may be proved by means of simple substitution for these quantities. For we have (Fig. 11)

$$dm' = 2\epsilon\psi\, d\psi,$$
$$dm = 2\epsilon r[\pi_{☾}/(\pi_{☾} + \pi_{\odot})]^2\, dr,$$
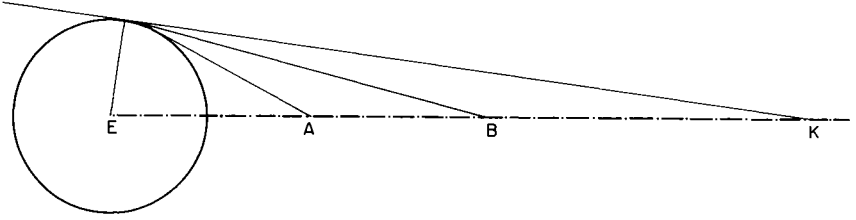
Fig. 11. Photometrical conditions in the cislunar space. A shows the limit of the astronomical and B the limit of the civil twilight; K represents the sunrise.

so that the ratio

$$\frac{dm}{dm'} = \frac{r\, dr}{\psi\, d\psi}\left(\frac{\pi_{☾}}{\pi_{☾} + \pi_{\odot}}\right)^2,$$

by substituting, becomes

$$\frac{dm}{dm'} = \left[1 - \frac{\omega}{\pi_{\odot} + \pi_{☾}}\left(1 - \frac{h_0'}{a}\right)\right]\left[1 - a\frac{d\omega}{dh_0'}\frac{1}{\pi_{\odot} + \pi_{☾}}\right] \tag{17}$$

which is identical with the expression (17) for the attenuation by refraction.

The description of these phenomena is no longer in the domain of idle theory, as it was at the end of the last century, when Proctor and Raynard (1892) dealt with these picturesque phenomena in some detail. The development of astronautics enabled us to observe the eclipse 1967 IV 14 from the Moon (Surveyor III, 1967) and transmit by television several interesting images of the eclipsed Sun. We can see on these, in addition to the refraction patterns, largely saturated in

the intensity, much weaker scattering halo interrupted at many angles by clouds (Section XIV).

Eclipse observation from orbiting spacecraft was recently examined by Moore and Schilling (1965). They first consider from the geometrical viewpoint the course of phenomena when the observer is leaving the Earth along the shadow axis. At the beginning of his travel, the observer will be plunged into the astronomical night until he reaches the point A (Fig. 11) where the first gleam of the astronomical twilight could be seen on the terrestrial limb. Going further, our passenger traverses successively the zone corresponding to the conventional stages of the twilight, i.e., astronomical, nautical, and civil. At the point K at the vertex of the refracted shadow cone, the Sun rises for him as a bright ring around the Earth. The distance of the above mentioned points can be easily calculated, taking into the account the refraction and the angular radius of the Sun. We have in this way the altitudes in the tabulation.

	Astronomical twilight	Nautical twilight	Civil twilight	Sunrise
Without refraction	13974	23648	52077	1,377,000 km
With refraction	13660	22948	49426	255,000 km

For the photometrical side of the phenomena, the authors have obtained a partial solution based on a semi-empirical method that combines refraction and multiple-scattering theory with observed measurements of the brightness of the twilight sky (Koomen *et al.*, 1952; Richardson and Hulbert, 1949). Their curve
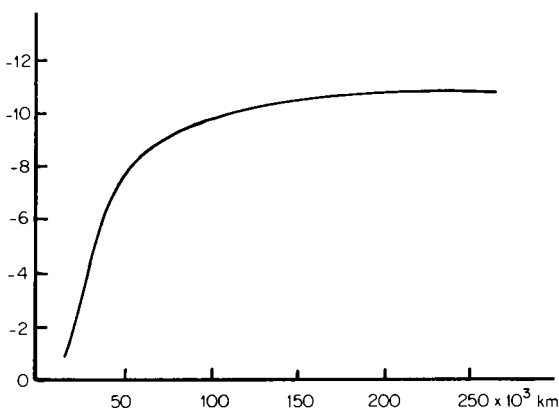
FIG. 12. Stellar magnitude of the atmospheric halo at different distances from the Earth (from Schilling and Moore, 1966).

gives the total visual brightness of the atmospheric halo as a function of the altitude (Fig. 12). Extrapolating this curve to the Moon, we obtain for the ratio of atmospheric to solar illumination outside the eclipse 3×10^{-7} instead of 2×10^{-7} obtained from (53), taking into the account only the primary scattering. However this new value can play a role only in the central parts of the shadow; in external parts of the umbra and in the penumbra, its influence is almost negligible.

Incidentally, similar phenomena, though on a different scale (and, unfortunately, only very rarely), may be observed during the transits of Venus across the Sun. Such transits may be used with advantage in the research of the physical conditions prevailing on Venus. The results of the observations from the latest four transits (1761, 1769, 1874, 1882) permitted an estimate of the air-pressure on the level of the cloud-layer and the determination of the temperature-variations along the limb of Venus. They simultaneously helped to refute the longstanding opinion concerning the equality of the period of rotation and revolution around the Sun (Link, 1959).

XIII

GEOGRAPHIC CIRCUMSTANCES OF ECLIPSES

For the purpose of a more detailed discussion of lunar eclipses, the position of the terminator of the Earth's shadow must be established (Link, 1947a). The terminator of the shadow is defined as the locus of points on the Earth's surface at which—with regard to refraction—the upper limb of the Sun rises or sets at the same time. Its position may be easily determined from astronomical ephemerides, where such moments are given for various days of the year and for various geographical latitudes. These times indicate simultaneously the geographical longitude of the terminator for 0 hr GMT positive eastward of Greenwich. At the time H of GMT, the terminator will be shifted for H westward. In this way, the terminator position may be plotted for arbitrary phases of the eclipse.

In general, of the entire terminator 40,000 km long, only a certain part participates in the projection of the shadow, which becomes apparent from the solar eclipse as seen from the Moon. This part is a section of the central angle ΔP where (Fig. 13)

$$\sin \Delta P = \frac{R_\odot'}{\gamma'} = \frac{R_\odot}{\gamma}. \tag{42}$$

If P is the angle of position of the point in question of the lunar plane measured from the north, the point half-way between in the arc of the terminator will have the geographical latitude

$$\sin \varphi = \cos \delta_\odot \cos P, \tag{43}$$
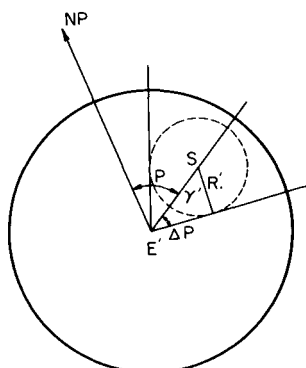
FIG. 13. Central angle of the shadow terminator.

and the end points of the arc will be characterized by the latitudes

$$\sin \varphi_{1,2} = \cos \delta_{\odot} \cos(P \pm \Delta P). \tag{44}$$

The refraction image of the Sun observed from the Moon is doubled. The second, lower and consequently fainter, image of the Sun corresponds to the arc of the terminator symmetrically conjugate with respect to the Earth's center with the first arc.
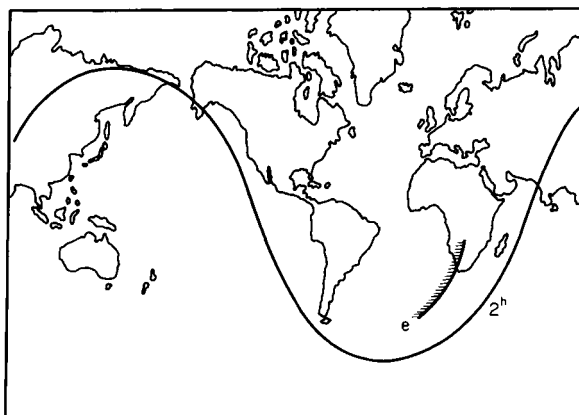
FIG. 14. Shadow terminator at the eclipse 1954 I 19 at 2^h GMT Position *e* corresponds to the end of the eclipse.

XIV

EFFECTS OF LIGHT SCATTERING

Scattering of light in the terrestrial atmosphere brings about a reduction of the intensity of the ray passing through it as well as a restitution of part of the lost radiation in the illumination of the shadow. In other words, the illuminated atmosphere becomes a secondary source of light which is added to the direct illumination of the shadow, as was first pointed out by Rougier and Dubois (1944). While the direct illumination decreases in the direction towards the center of the shadow, the parasitic light arising from diffusion changes only slightly, and might have an appreciable effect in the central parts of the shadow.

The brightness of the elementary volume db of the diffusing medium is given by the expression (Link, 1950a),

$$db = CE\,dD, \tag{45}$$

where E is the illumination, and dD the optical density in the direction of the ray. Let A (Fig. 15) be the entrance and B the exit points in the ray passing through the scattering medium, and K an arbitrary point between A and B. The brightness of a volume element in point K will be

$$\begin{aligned} db &= CE_0 \exp(-D_1 - D_2)\,dD = CE_0 \exp(-D_0)\,dD, \\ D_0 &= D_1 + D_2\,, \end{aligned} \tag{46}$$

where D_1 and D_2 are the optical densities between BK and AK.
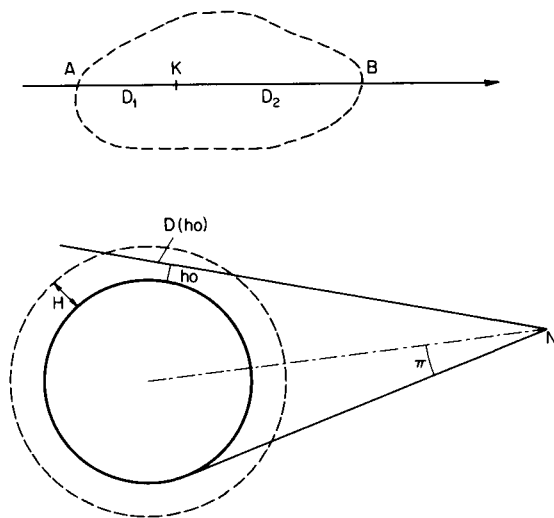
FIG. 15. Light scattering in the atmosphere.

Let us apply this formula to the ring of terrestrial atmosphere of a radius of $a + h_0$ and of width dh_0. Its brightness will be

$$b(h_0) = \int db = CE_0D(h_0) \exp[-D(h_0)] \tag{47}$$

and, viewed from the Moon, it will appear under the solid angle

$$d\Omega = \frac{2\pi[1 + (h_0/a)^2] \, \pi_{☾}^2}{a} dh_0; \tag{48}$$

so that the illumination on the Moon will be

$$d\eta = b(h_0) \, d\Omega \tag{49}$$

and the illumination by the whole atmosphere

$$\eta = \frac{2\pi CE_0\pi_{☾}^2}{a} \int_0^H \left(1 + \frac{h_0}{a}\right) D(h_0) \exp[-D(h_0)] \, dh_0, \tag{50}$$

where we have integrated over the whole height H of the atmosphere.

The upper limit of this illumination will be

$$\eta < \frac{2\pi CE_0\pi_{☾}^2}{a} 0.368 \left(H + \frac{H^2}{a}\right) \tag{51}$$

or at a mean parallax of the Moon, $\pi_{☾} = 57'$ and $C = 1/4\pi$

$$(\eta/E_0) < 7.910^{-9}H + 1.210^{-12}H^2, \tag{52}$$

where we have assumed an orthotropic diffusion of light. For $H = 100$ km, we obtain $\eta/E_0 = 8 \times 10^{-7}$. Švestka (1948) computed the case of diffusion for the actual atmosphere by numerical integration, and obtained

$$(\eta/E_0) = 1.9 \times 10^{-7} \quad (7000 \text{ Å}) \qquad \text{or} \qquad 2.1 \times 10^{-7} \quad (4500 \text{ Å}). \tag{53}$$

These estimates of the upper limit of parasitic light will be used further on in Section XX and XXI.

XV

PHOTOMETRY OF LUNAR ECLIPSES

The aim of the photometry of lunar eclipses is to determine the shadow-density by the expression

$$D = \log_{10}(E/e). \tag{14}$$

Up to the present, this determination has been attempted, with varying success, by the visual, photographic, and photoelectric methods.

All measurements of the shadow-density have to recognize several difficulties inherent in the nature of the measured object. During the partial phase, there is a great light contrast (1 : 100) between the part of the Moon in the penumbra and the part in full shadow. Parasitic light from the bright crescent of the Moon diffused within the atmosphere and on the optics of the instrument penetrates into the pencil of light from the full shadow, and reduces its apparent density. Moreover, further trouble arises from the variation of extinction during the eclipse, since in its course we relate the initial measurement within the penumbra to the subsequent measurement within the shadow. Finally, certain difficulties may result from the great amplitude of the intensities in the ratio 1 : 10^4 and more, and the associated faint intensity in the full shadow. Therefore, in addition to the required spectral selectivity of the radiation receivers, also a high sensitivity becomes indispensable. The efficiency of the photographic method is further limited by the rapid changes during the eclipse which do not permit longer exposures.

The most reliable and at the same time most numerous results have been obtained up to now by Danjon's visual method (Danjon, 1928) utilizing his cat's-eye photometer. In principle, this is a double-image photometer comparing the brightest and faintest parts of the lunar disk in the course of the eclipse. In front of the objective O (Fig. 16), there is a system of two prisms P_1 and P_2,
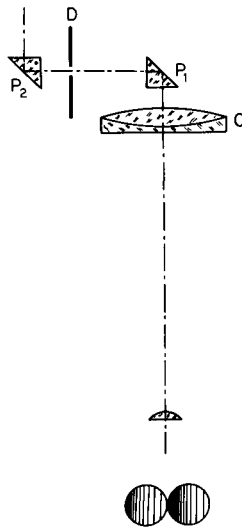
FIG. 16. Danjon's photometer with cat's eye. Below, view of the Moon during measurement.

which only partially covers its free aperture. In the focal plane of the objective two images of the Moon may be observed—namely, a direct image produced by the free aperture of the objective, and an indirect image produced by reflection on the system of prisms. By means of their variable orientation, we easily obtain an external contact of the brightest and faintest parts of both images (Fig. 16) during the eclipse. It is, therefore, necessary to tilt the system of prisms also tiltable around the optical axis of the objective into the position angle of the line connecting of the centers of the Moon and shadow. A square diaphragm D is inserted in the optical path of the rays passing through the prisms, the aperture of which is continuously varied (cat's-eye). My means of this diaphragm, the observer equalizes the brightnesses of both images where they come in contact.

During the eclipse, the measurements commence when the Moon enters the penumbra, and give the density curve in the penumbra. The same procedure is adopted when the Moon begins to enter into the shadow, whereby the density-curve in the edge part of the shadow is obtained—related, of course, to the penumbra, the profile of which has already been ascertained at the beginning, and, in the same manner, we proceed during the total eclipse, and its second partial phase. These results are, of course, affected by the differences of the albedo along the Moon's limb, which must be determined by additional measurements before or after the eclipse. The parasitic light in the field of view of the photometer is relatively harmless, as it disturbs practically in the same way both adjacent plages. The extinction is usually also harmless, since it appears only differentially as a difference of the extinctions between both limbs of the Moon. The sensitivity of the eye is sufficient to permit the use of three-color filters as far as the center of the shadow.

For a successful performance, Danjon's method requires of course a longer series of measurements which would enable a continuous intercomparison of the individual parts of the penumbra with the full shadow. Furthermore, it is tacitly assumed that the density depends only on the distance γ, and not on the position angle P. Whenever the Moon does not pass through the shadow radially—which is a very frequent case—this may cause a certain heterogeneity of the results.

Of the photographical methods specially adapted to the purposes of lunar eclipses, let us mention that of Link and Guth (1936) which, to a certain extent, eliminates the above-mentioned disturbing influences. Immediately in front of the photographic plate on which the image of the Moon is projected, we place a neutral gelatine filter F (Fig. 17) of the density of about 2, cut to the radius of the shadow. The filter is inserted so as to cover precisely the part of the Moon in the penumbra which, to a certain degree, compensates for the great light contrast between the shadow and the penumbra. The scattered parasitic light will be determined from the intensity of the fog in the vicinity of the Moon's image. Moreover, by means of an incandescent lamp L, an intensity scale will be impressed on the plate. Otherwise the procedure of the measurement—that
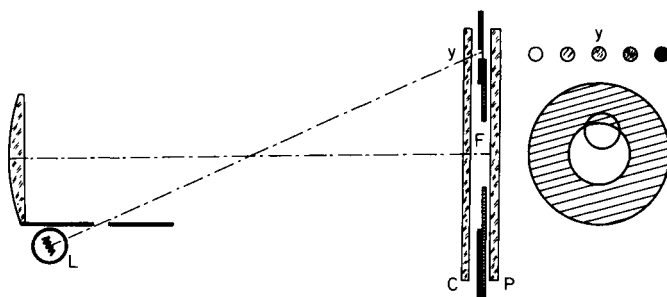
FIG. 17. Link's photographic photometer. C–color filter, F–neutral filter, I–intensity scale, P–photographic plate.

is, a continuous intercomparison of the individual parts of the penumbra and shadow—is similar to that of Danjon's method, whereby also the influence of the extinction is eliminated. An advantage of this method consists in the possibility of determining the isophotes of the shadow in those parts which the Moon's disk has traversed. The selectivity of the filters and, consequently, also the sensitivity of the method are limited by the rate of the movement of the shadow which does not permit longer exposures.

The photoelectric method also has been adapted to the purposes of lunar eclipse. Walker and Reaves (1957), in order to reduce the effect of parasitic light, measured only that part of the Moon which lies farthest from the center of the shadow. This part is always the brightest and, consequently, the influence of the parasitic light is relatively smallest. Cimino and Fresa (1958) photoelectrically compared two opposite points of the Moon, and by this practically eliminated the influence of extinction.

XVI

COMPARISON BETWEEN THEORY AND OBSERVATIONS

This comparison will be based on the differences $\Delta = O - C$ between the measured and computed density of the shadow or penumbra, at various angular distances γ from their center. Under application of his theory and on the basis of observational material which, as time went on, became more abundant, Link (1956b) carried out detailed comparisons of this kind. In his material, the prevalent measurements were those made by Danjon and his followers, Rougier and Dubois.

The final results may be summarized as follows. In the penumbra, near to the edge of the shadow (Fig. 18), the observed shadow-density is usually lower than that obtained by computation. Furthermore, the results—both measured
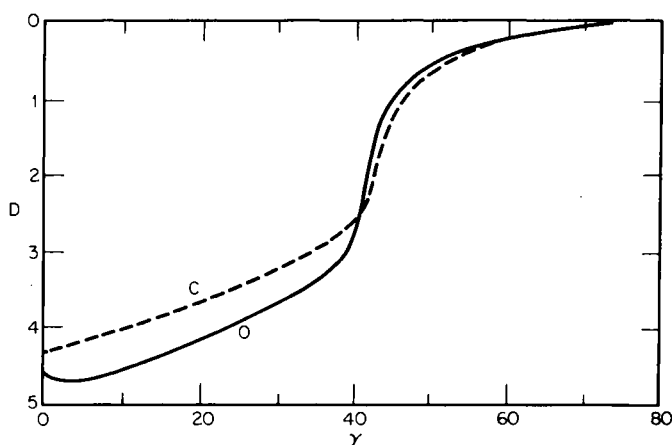
FIG. 18. Schematic course of the density of the umbra and penumbra. O–observed, C–computed density curve.

and computed—are practically independent of the color of the light. Thus, in this part, the penumbra is brighter than suggested by the theory, in all colors. Close to the edge and within the shadow, the densities obtained by measurements are mostly greater than the computed ones. These differences are greatest in the red color and diminish towards the violet end of the spectrum. In the blue, they frequently happen to be negative (i.e., the measured density comes out lower than the computed). The magnitude of the positive values of Δ (in the green and red colors) does not vary too much across the disk; and in the blue color, the amount of these differences even decreases towards the center of the shadow.

In addition to this variation of the differences Δ with the angular distance γ from the center, attention must be given also to their dependence on the angle of position at a constant distance from the center of the shadow. The position angle from the equator is $P' \approx \varphi$, i.e., P' is equal to the geographic latitude of the effective part of the terminator (Section XIII). This dependence of Δ on P' or φ, changes its sense on both sides of the shadow's edge. In the penumbra, Δ increases in the direction towards the equator, while in the shadow it usually decreases in this direction. All these peculiarities and some others should now be the object of our analysis.

XVII

ATMOSPHERIC OZONE

First, we shall attempt to explain the behavior of the differences Δ in the central and outer parts of the shadow. In Fig. 19 a typical case observed during the
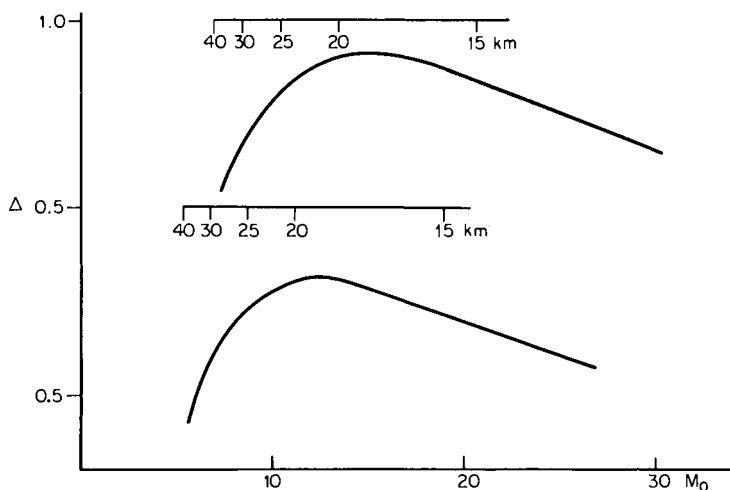
FIG. 19. Curve of the differences $\Delta = O - C$ depending on the mean air-mass M_0 for the eclipse of 1921 X 16. Scale of the altitude of the ray h_0 corresponds to the upper limit of the illumination integral.

eclipse of October 16, 1921, is shown, which has been measured by Danjon (1921) in the green and red regions of the spectrum. In Fig. 19, the differences Δ are plotted as functions of the mean airmass. Moreover, to certain points of the curve, the minimum elevations h_0 of the rays are added, corresponding to the upper limit of the integral of illumination (12)—i.e., of the highest rays from the whole pencil illuminating the measured spot within the shadow.

The trend of the differences Δ may be explained by the absorption of light in the terrestrial ozone layer. Let us assume, in the first approximation, that the ozone is concentrated in a thin layer between the altitudes h_1 and h_2 (Fig. 20). The trajectory of a horizontal ray of altitude h_0 in the layer will, if the refraction is neglected, be

$$\frac{G(h_0)}{G(0)} = \frac{(h_2 - h_0)^{1/2} - (h_1 - h_0)^{1/2}}{h_2^{1/2} - h_1^{1/2}}, \tag{54}$$

where $G(0)$ is the trajectory of the ray of the zero altitude $h_0 = 0$.

The ratio (54) rises at first only very slowly, but later more rapidly, to the maximum attained at the lower edge of the layer, only to drop steeply almost to zero at the upper edge of the layer. A similar behavior may also be observed in the differences which begin rapidly to diminish below the altitude of about 21 km. This is the level at which the lower limit of the ozone layer must be located. Earlier, in fact, the ozone layer was placed much higher, at about 50 km (Cabannes and Dufay, 1927), and only the first results from lunar eclipses
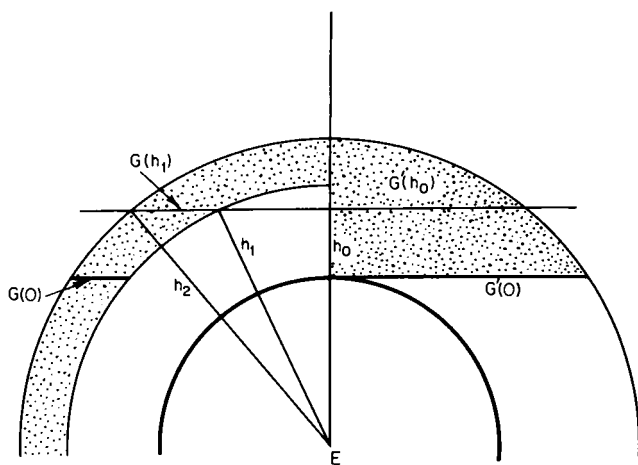
FIG. 20. High absorbing layers. Left–thin, right–thick layer.

(Link, 1933) directly indicated its lower position which also had been indirectly derived by means of the Umkehreffekt (Götz, 1931).

Similar curves of the differences were derived also for other eclipses (i.e., for other positions of the terminator on the Earth). Here, essential differences depending on the geographical latitude of the terminator became apparent. In Fig. 21 some such curves (Link, 1946a) have been plotted. The curves from lower
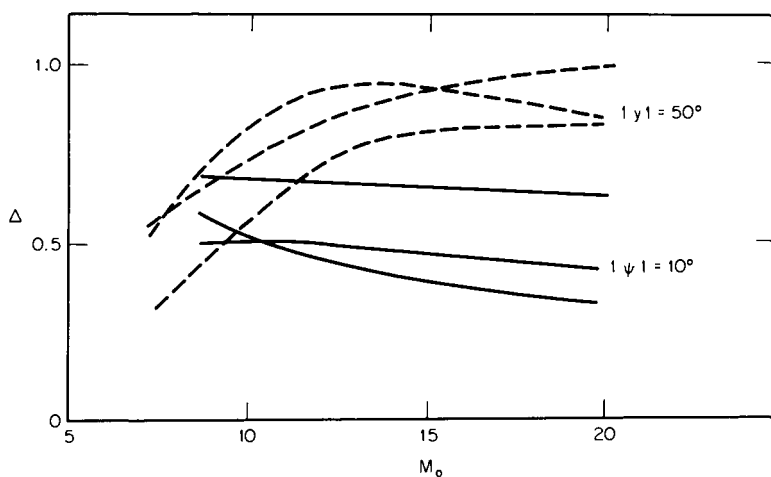
FIG. 21. Curves of the differences Δ for several eclipses from lower and higher latitudes as functions of mean air-mass M_0.

latitudes (10°) give smaller differences than those from higher latitudes (50°). Accordingly, the amount of ozone in the atmosphere seems to decrease towards the equator, a fact which has been confirmed by direct measurements (Dobson, 1930). In the curves from higher latitudes, the steep drop of the differences is well noticeable, while in the curves from lower latitudes this drop is not apparent within the limits of our diagram. This would seem to indicate that the altitude of the ozone layer increases towards the equator, and Dobson's measurements in Arosa and Tromsö confirm this phenomenon. In other words, the isophotes of the central parts of the shadow are flattened on the equator and elongated on the poles, owing to the varying amount of terrestrial ozone in various geographic latitudes.

A more accurate determination of the distribution of ozone with the altitude during lunar eclipses was attempted by Paetzold (1950, 1951, 1952). When the refraction is neglected, the total amount of ozone along the ray passing through at the minimum elevation of h_0 may be expressed (cf. Fig. 20) as

$$O(h_0) = (2a)^{1/2} \int_{h_0}^{\infty} \frac{o(h)\, dh}{(h - h_0)^{1/2}}, \tag{55}$$

where $o(h)$, the concentration of ozone at the altitude h, is known from spectrophotometric measurements carried out by the author. From this equation, the function $o(h)$ must be determined. This can be done by recognizing in (55) an integral equation of Abel's type, whose inversion yields

$$o(h) = -\frac{1}{\pi(2a)^{1/2}} \int_{h}^{\infty} \frac{[dO(h_0)]/(dh_0)}{(h_0 - h)^{1/2}}\, dh_0 . \tag{56}$$

The observations do not, of course, give the amount of ozone for a specific ray of the altitude h_0, but for the whole pencil of rays illuminating the given point in the shadow—i.e., they provide us with some kind of the mean value $O^*(h_0)$. This difficulty may be obviated by substituting for the whole pencil a mean ray passing through the air-mass equal to that of M_0 (Section XI). From spectrophotometric measurements of several eclipses, Paetzold (1952) obtained rather good curves of the function $o(h)$ of the distribution of ozone with the altitude (Fig. 22). Similar results were reported by Vigroux (1954) from the eclipse of January 29, 1953 (Fig. 20), when on the equator the possibility of a double maximum was indicated. On the north pole, the concentration of ozone is higher, and the maximum of the curve lies lower than on the equator. This agrees well with what has been known of the ozone layer so far.

Today, however, when the distribution of the atmospheric ozone is relatively well known from balloon and rocket ascents, the whole procedure may be reversed. From the known distribution of $o(h)$, and from the absorption coefficients of ozone obtained by laboratory experiments, we can compute its optical
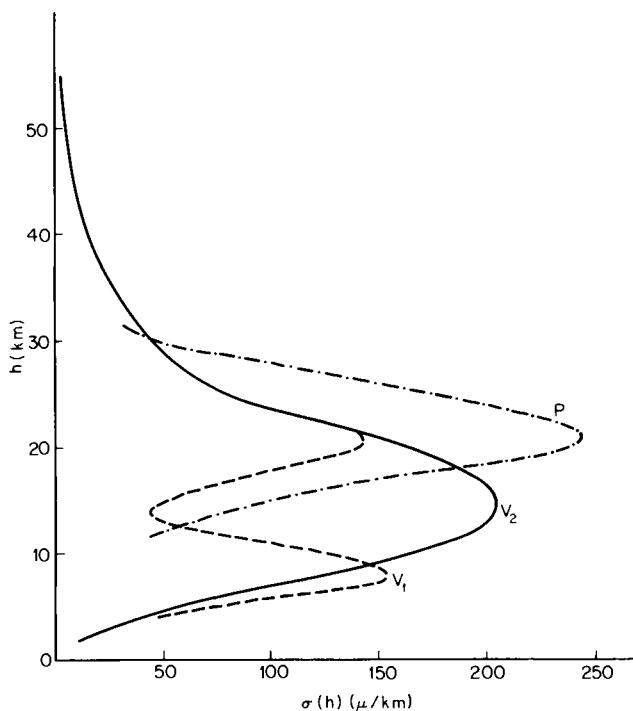
FIG. 22. Curves of ozone distribution, according to results of eclipses V_1 (equatorial), V_2 (polar), Vigroux, 29th January, 1953, Paetzold, 16th October, 1921 (measured by Danjon, Fig. 17).

density for various altitudes of horizontal rays. In the first approximation, we proceed according to Eq. (55), or we may take into account the refraction. The element of ozone mass is then

$$dO = o(h) \sec i \, dh$$

and the total mass

$$O = 2 \int_{h_0}^{\infty} o(h) \sec i \, dh.$$

Having introduced the new variable $Z = \cos i$, we get

$$O = 2 \int o(h) N \, dZ \qquad \text{with} \qquad N = 1/(1 - c\rho\beta). \tag{56'}$$

The integration may be achieved by numerical methods. Table VIII summarizes the results of such computations based on the mean values of function $o(h)$.

The first computation is by Link and Guth (1940) with the mean distribution according to the terrestrial and rocket measurements. The second computation, by Hansen and Matsushima (1966), is based on the empirical formula of Green (1964)

$$o(h) = \frac{P_1 \exp P(h)}{P_2[1 + \exp P(h)]^2}, \tag{57}$$

where

$$P(h) = (h - P_3)/P_2$$

and P_1, P_2, P_3 are parameters adjusted to fit the best distribution after Green.

TABLE VIII

h_0	5	10	15	20	25	30	35	40 km
				Link and Guth (1940)				
$o(h_0)$	4.1	5.4	7.2	9.1	10.3	9.1	4.5	1.5×10^{-3}
$O(h_0)$	8.9	9.1	9.0	8.7	7.8	4.7	1.9	0.3 cm
				Hansen and Matsushima (1966)				
$O(h_0)$	7.0	8.1	9.3	9.4	7.0	3.7	1.5	0.6 cm

In the zenith, the total amounts of ozone are 0.27 cm or 0.26 cm, respectively. As can be seen, the total amount of ozone $O(h_0)$ is somewhat variable at altitudes ranging from 0 to 20 km, but equals approximately thirty-three times the amount of ozone in the zenith. This means that in the auxiliary shadow (Table III) as far as 52′, and in the actual shadow as far as to $52' - 16' = 36'$, the absorption by ozone is nearly constant. Its value for various wavelengths obtained from the known laboratory values of the absorption coefficient have been computed in Table IX. Under actual field conditions, the annual and geographic variations of ozone should be taken into consideration. Their amplitude, however, will not surpass 100% of the value given in the table (Mitra, 1952).

TABLE IX

CONSTANT OZONE ABSORPTION

λ	4600	5400	5800	6200	6800	7000	7600 Å
Δ	0.03	0.33	0.53	0.43	0.15	0.09	0.02

XVIII

HIGH ABSORBING LAYER

A comparison of the absorption by ozone with the differences $\Delta = O - C$ found by measurements indicates that, in most cases, the ozone alone cannot explain them adequately. We still encounter remainders attaining 0.2 to 0.7 in densities which are also constant in a wide range of the air-masses, and which cannot be explained by ozone absorption (Fig. 23). We associate it with a high absorbing
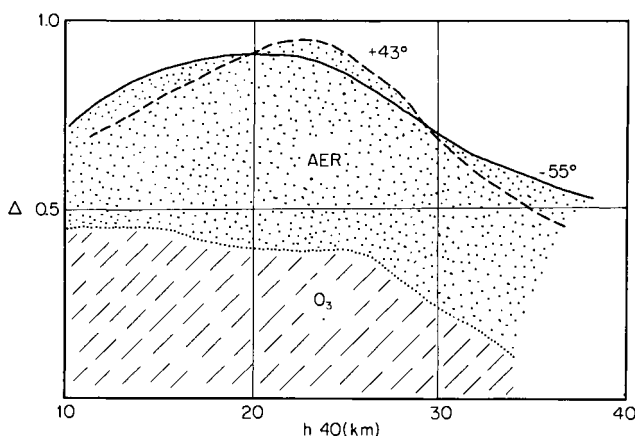
FIG. 23. Remainders $O - C$ for two different eclipses at 43°N and 55°S. Aer–the absorption of aerosols, O–the absorption by ozone; on the horizontal axis—the upper limit of the illumination integral (from Link, 1960b).

layer. Unlike the case of the ozone layer, its upper limit seems to lie much higher —around 100 km—and, for lack of more detailed information, we shall consider it tentatively as homogeneous. The ratio of the trajectory of rays is given, analogously with Eq. (54), by the formula

$$\frac{G'(h_0)}{G'(0)} = \left(1 - \frac{h_0}{h_2}\right)^{1/2}. \tag{58}$$

Its course may be best seen from the numerical values of the density of the layer in Table III, computed for a density of 0.006 at the zenith, and with an upper limit of the layer $h = 100$ km. It is constant for a considerable range of altitudes, and drops steeply only in close vicinity of the upper limit, as can be seen in Fig. 24.

The high-absorbing layer reveals itself in the phenomena of eclipses, particularly by an enlargement of the Earth's shadow. A treatment of all previous determinations (Link and Linková, 1954b) indicates an average value of 2.3%. The
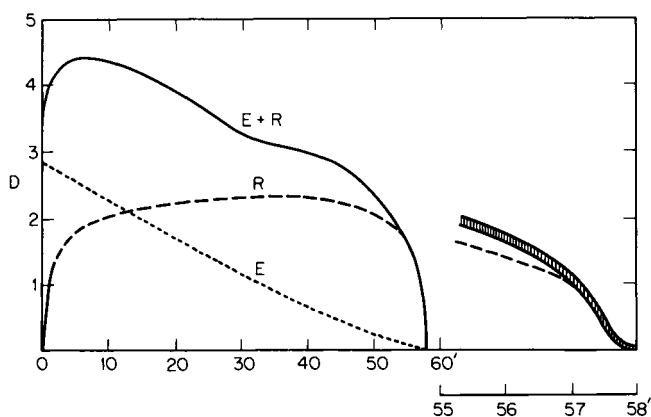
FIG. 24. Influence of the high absorbing layer on the density of the auxiliary shadow. E–extinction, R–attenuation by refraction; dashed line-influence of the absorbing layer. Detail is shown at right.

enlargement of the shadow is a physiological phenomenon. At the edge of the shadow, where illumination e changes very rapidly with the distance γ, the eye locates the edge of the shadow by estimating the point of the fastest change—i.e., the inflection point of the curve $D = f(\gamma)$ (Kühl, 1928). This situation was analyzed experimentally by Paetzold (1953), who repeated Seeliger's experiment (1896) with a rotating sector (Fig. 25), the central angle of which changes with the radius so as to be in every point proportional to the illumination in the corresponding point of the shadow. If rotating rapidly, the sector appears as a disk with a black central portion and a hazy border. The observer's task consists in determining, by means of a simple measuring instrument, the apparent radius of the disk. Experiments with various forms of sectors computed for
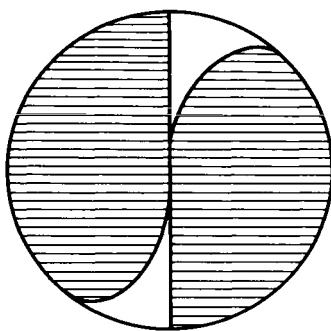
FIG. 25. Seeliger's sector.

various shadow structures revealed that only the high-absorbing layer may bring about an enlargement of the shadow; neither ozone nor a low-lying layer (like, for instance, fog) are effective to this end. This circumstance is also evident from the course of the density of the auxiliary shadow. The most rapid change in illumination will apparently occur at the moment when the Sun emerges from the comparatively sharp border of the auxiliary shadow (Fig. 24) which is apparently enlarged by the high-absorbing layer.

As has been shown by Kosik (1940), the enlargement of the shadow depends on the angle of position P'. The limit of the shadow may be considered as an ellipse which, however, is more flattened than may be computed from the oblateness of the geoid. For the eclipse, for instance, of May 3, 1939, Kosik found

$42.90' - 0.456' \sin^2 P'$	observation
$42.12' - 0.183' \sin^2 P'$	computation
$0.78' - 0.273' \sin^2 P'$	enlargement

His results were also confirmed by later researches (Bouška, 1948; Koebke, 1951), the most important of which is the extensive treatment of the results from the years 1889 to 1938 (Link and Linková, 1954a). A similar phenomenon may be observed on the border of the shadow, where it appears in the form of the curve $D = f(\gamma)$. Curves from lower latitudes (Fig. 26) exhibit a prominent bending, while curves from higher latitudes are more rounded.

The flattening of the shadow is also supported by photometric measurements of the density of the penumbra in the vicinity of the edge of the shadow. From a treatment of the measurements secured by means of the cat's-eye photometer, Link (1959) found that, whenever the measurements had been made in the first and second parts of the eclipse, greater densities were registered in the equatorial parts of the penumbra than in higher latitudes; since the geographic latitude of the effective part of the terminator of the shadow was different in each part of the eclipse. This fact is even more apparent in the results of photographic photometry of the eclipse of December 19, 1945, where the isophotes of the interior penumbra (Link, 1948b) show a flattening towards the poles.

The explanation of the flattening of the shadow must be sought in the variable structure of the terrestrial atmosphere around the altitude of 100 km, varying with geographical latitude. At these altitudes, it is absolutely impossible to take into account the extinction by ozone, or an attenuation of the refraction (Table III); so that the only factor known up till now might be the high absorbing layer—composed, in all probability, of dust particles—which would be more elevated and dense in the equatorial regions than on the poles.

If, as has been mentioned above, the value of absorption for horizontal rays amounts to 0.2 to 0.7 in density of the shadow, it certainly should be much lower in the radial direction, in which it may be determined from terrestrial
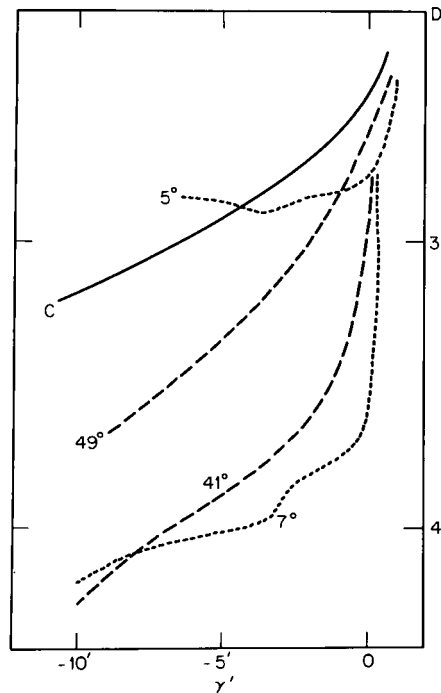
FIG. 26. Curves of density of the shadow in the vicinity of its geometrical limit Figures indicate the mean geographic latitude of the terminator; C denotes the computed curve.

observations. The ratio of both absorptions depends on the structure of the layer. If it is homogeneous—an assumption which may be rather far from the actual reality—it should be approximately 1 : 20. Actually, it will be higher, if we admit an increase in the density of the layer with decreasing altitude. The fact that such a medium exists in the high atmosphere had already been demonstrated before its identification during eclipses (Link, 1933) by other proofs—however indirect or unconvincing they may have seemed.

First of all, attention has to be paid to the measurements of the atmospheric absorption by means of Bouguer's straight lines. The logarithm of brightness $\log I$ of the extraterrestrial body (e.g., the Sun) is measured at various zenith distances z and plotted in the graph as a function of the airmass $M(z)$ (Fig. 27). The points lie approximately along the straight line, the inclination of which gives the requisite absorption coefficient. Two effects transpire, which may be explained by the presence of a high absorbing layer. Actually the points do not lie along a straight line, but rather along the curve c. At small zenith distances,
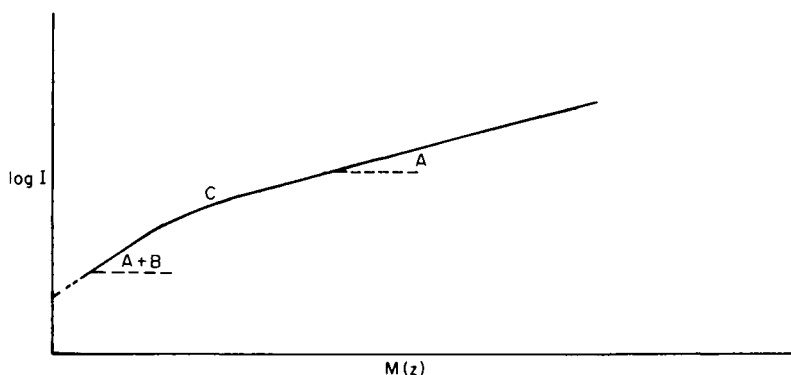
FIG. 27. Bouguer's graph (from Link, 1963).

when the path in the layer is changing with the air-mass $M(z) \sim \sec z$, the following equation may be used

$$\log I = \log I_0 - (A + B) \sec z, \tag{59}$$

in which B is the absorption coefficient of the layer (i.e., its optical density at the zenith).

Close to the horizon the air-mass $M(z)$ changes very quickly with the zenith distance, whereas the path in the layer is practically constant and equals $G(0) = (2a/h_2)^{1/2}$; the equation of this part of the curve is then

$$\log I = \log I_0 - AM(z) - B(2a/h_2)^{1/2}. \tag{60}$$

The occurrence of the curvature is thus an effect pointing to the possibility of a high absorbing layer. Such an effect was found at first in Müller's measurements at Potsdam (1883); and Hausdorff (1895) was the first to offer the explanation by means of a high absorbing layer. Later on similar results were arrived at by Bauer and Danjon (1923) on the basis of measurements taken on Mt. Blanc, and by Link on Pic-du-Midi and elsewhere (Link, 1929). The results give the value of absorption at the zenith $B = 0.015$–0.045 and the altitude of the layer above 100 km. The use of the method is rather delicate, for the construction of Bouguer's straight line or curve requires very stable conditions in the atmosphere, which seldom occur even at high mountains.

Another method leading to the hypothesis of the high absorbing layer was put forward by Linke (1943), Götz (1940), and Link (1943). It is based upon the formula given by Rayleigh and corrected by Cabannes (1929) for the absorption coefficient A of the ideal atmosphere

$$A = 4.57 \times 10^{-12}(c^2/\lambda^4),$$

or in logarithmic form,

$$\log A = \text{constant} + \log(c^2/\lambda^4). \tag{60'}$$

If we represent now the dependence of $\log(A + B)$, we obtain a curve that differs from the line inclined by 45°, as should be true for the coefficient A alone (Fig. 28). In the region of 6000 Å, the deviation from it is caused by ozone absorption. Otherwise, the differences between the observed and theoretical curves (Fig. 28) are practically independent of the wavelength and of the altitude of the observation station—particularly if we confine ourselves to the series of
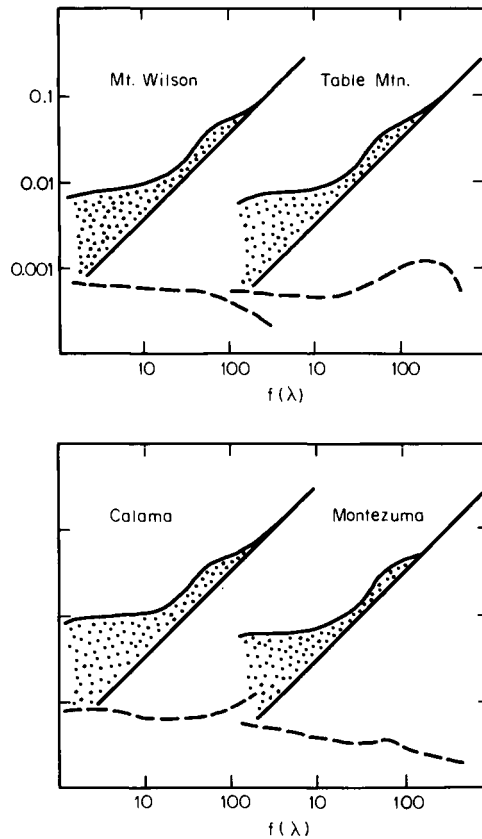
Fig. 28. Linke's diagram. Dependence of the logarithm of optical density in the zenith on $\log c^2/\lambda^4$. The inclined line gives the theoretical relation (60) for Rayleigh's atmosphere. Below: dependence of logarithms $O - C$, corresponding to the dashed region above with correction for ozone absorption in the orange region of the spectrum. The curves are valid for days of perfect transparency at the mentioned stations.

measurements secured at the times of greatest atmospheric transparency. This difference ranges around 0.006 in density, and we attribute it to the high absorbing layer.

A few words remain to be said on the probable origin of this high absorbing layer. The theory of its meteoric origin seems very plausible. According to various phenomena reported in other branches of our science, it appears probable that a certain amount of interplanetary matter penetrates into the high atmosphere. Larger particles evaporate at altitudes of about 100 km; and smaller particles—micrometeorites—are only slowed down by the resistance of the gaseous medium. All such particles then fall freely through the molecular medium of the Earth's atmosphere. A simplified theory of this phenomenon (Link, 1950c) considered particles of equal radius as well as of equal density. If the total meteoritic accretion is m grams per cm^2 per second, and the duration of the fall of the particle is T, the optical density of the layer in the vertical direction should be

$$B = f(a)\, mT, \tag{61}$$

where $f(a)$ is the coefficient, computed by Greenstein (1937), depending on the semidiameter a and on the composition of the respective particles. The time of fall may be computed from the air density at various altitudes by means of the known formula for analogous cases occurring in the kinetic theory of gases.

The observational results show that the optical density of the layer in a vertical direction about 0.006 and that it is practically independent of wavelength. According to Greenstein's computations, this requires that $a > 3 \times 10^{-5}$ cm. The time of fall is, of course, not known directly, but certain phenomena indicate its order of magnitude. Thus Švestka (1950) found that the average brightness of the eclipse on Danjon's scale abruptly decreased following the maximum of meteor swarms (Fig. 29), attaining a minimum after 30 to 40 days, while its normal value is not regained till 80 to 100 days after the maximum. The
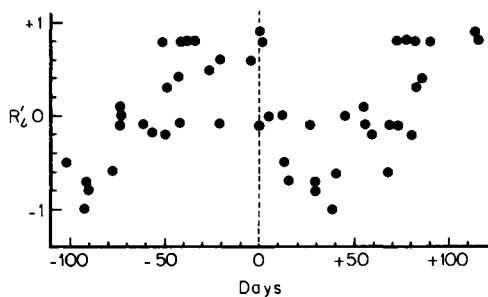
FIG. 29. Luminosity of lunar eclipse following the maximum of meteor swarms according to Švestka (1950).

period of 30 to 40 days would, therefore, seem to correspond to the time of fall of most of the micrometeoritic material. Bowen (1953, 1956, 1957) showed in several papers that, on the average, a sudden increase of precipitations may be observed 30 days following the maximum of meteoritic swarms; and he interprets this phenomenon by the penetration of meteoritic particles into the troposphere, where they become active as catalyzers of rain. Zacharov (1952) found that the transmission coefficients registered on Mt. Wilson are systematically lower for 24 days following the maximum of Perseids. Hansa and Zacharov (1958) found, moreover, in their dust collections lasting 14 days an increase of the ratio nickel : iron for intervals of time equal to the 2 and 3 multiples of the period of collection. All these phenomena taken together indicate an average time of fall in the neighborhood of 30 days; and the smallest micrometeorites which are also optically the most effective ones may need as long as 200 days to descend to the ground.

If we take for the upper limit the time of fall to be $T = 10^7$ sec, a diameter of the particles results as $a = 10^{-4}$ cm (Fig. 30) and the respective absorption-
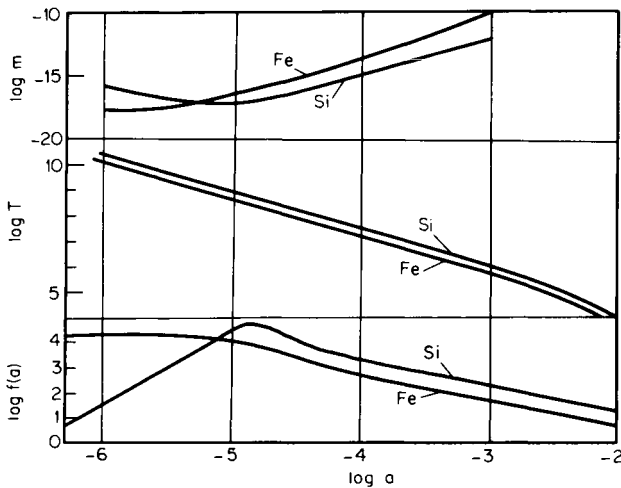
FIG. 30. Curve of the regime of the high absorbing layer. Above: meteoric accretion required for a density $\Delta = 0.004$ in the zenith; center: duration of the time of fall T of meteoritic particles; below: Greenstein's absorption coefficient $f(a)$, all as functions of the radius a of the particle and its composition; Fe–iron, Si–silicates.

coefficient $f(a) = 10^3$. From Eq. (61), we obtain, for a known amount of mass accretion m_0, the optical density in the zenith B and vice versa. The amount of meteoritic accretion (Table X) estimated from various phenomena, ranges between 10^{-11} to 10^{-14} gr/cm^2 · sec, which leads to an optical density at 0.01 to

TABLE X

METEORIC ACCRETION[a]

No.	Author	Year	Accretion (gm/cm² 5cc)	Method
1	Oppolzer	1884	5×10^{-10}	Upper limit according to the celestial mechanic
2	Hansa-Zacharov	1958	3×10^{-11}	Ni in terrestrial dust collections
3[b]	Petterson	1958	1×10^{-12}	Ni in terrestrial dust collections
4	Zacharov	1961	7×10^{-12}	Ni in terrestrial dust collections
5	Link	1955	5×10^{-13}	Optical density of the atmosphere
6	Petterson-Rotschi	1950	2×10^{-13}	Ni on the bottom of the Pacific
7	Petterson	1958	6×10^{-14}	As No. 3
8	Petterson	1960	4×10^{-14}	As No. 6
9	Petterson-Rotschi	1950	2×10^{-14}	As No. 6
10	Watson	1956	2×10^{-14}	All interplanetary matter, upper limit
11	Kreiken	1959	2×10^{-14}	Only magnetic spherules in collection
12	Kizilirmak	1960	2×10^{-14}	Magnetic particles in collections
13	Petterson	1960	2×10^{-14}	Air filtration, preliminary value
14	Thompsen	1953	1×10^{-14}	Magnetic spherules in collections
15	Van de Hulst	1947	4×10^{-15}	Zodiacal light
16	Watson	1956	2×10^{-15}	As No. 10, lower limit
17	Öpik	1956	2×10^{-15}	As No. 15
18	Norris-Hogg	1949	2×10^{-15}	Spherules
19[c]	Dubin	1960	1×10^{-15}	Space probes
20	Schwinner	1936	1×10^{-15}	Geological evidence, max. value
21	Schwinner	1936	3×10^{-16}	Geological evidence, min. value
22	Watson	1939	7×10^{-17}	Meteors and meteorites
23	Wylie	1935	2×10^{-17}	Meteors and meteorites
24	Watson	1939	7×10^{-19}	Only meteors

[a] From Zacharov (1961) and Link (1963).
[b] No. 3, the value of No. 7 reduced to the fall time of 30 days.
[c] No. 19, only a partial accretion, total accretion is 2×10^{-14}.

0.0001 at the zenith; while the rates of accretions estimated from observations of visual meteors from 10^{-17} to 10^{-19} are found by far inadequate to produce a measurable absorbing layer.

Very substantial is also the correlation between the enlargement of the shadow and the meteoritic activity as found by Bouška and Švestka (1950) in Hartmann's material, and as confirmed, later, from much more abundant material on 57 eclipses (Fig. 31).

In conclusion it may be said that the meteoritic origin of the high absorbing layer agrees well with other observed phenomena; although from the quantitative point of view this hypothesis requires more accurate data, for various parameters are concerned which are difficult to determine.
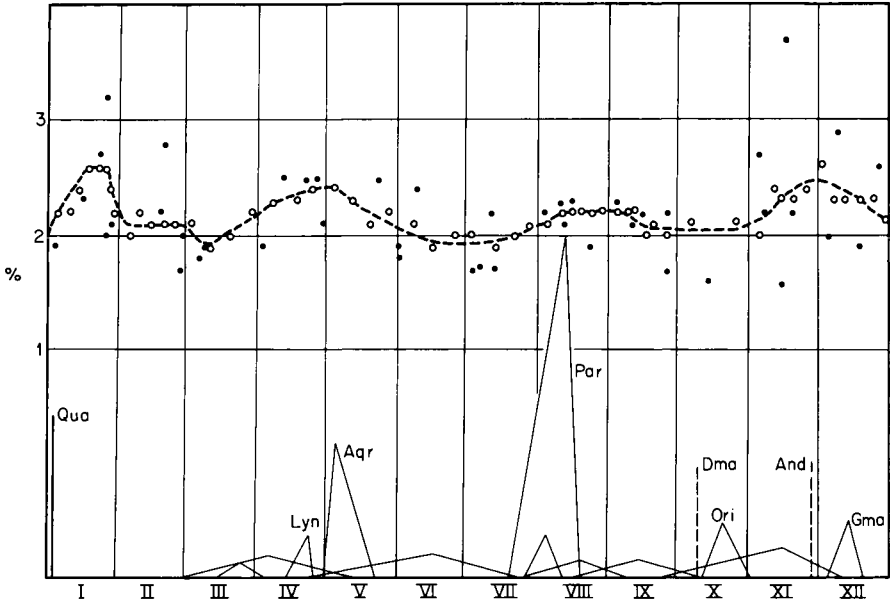
FIG. 31. Enlargement of the Earth's shadow as a function of the meteoritic activity (Link and Linková, 1954b).

As far as the lunar eclipses concerned, it is difficult to decide between the thin and the thick layer (Fig. 20). Measurements of the brightness of the sky at the zenith during the twilight will, however, prove useful (Link, 1948a). The brightness is due to the light scattering in the layers of the atmosphere at altitudes which depend on the actual height of solar rays. The absorption in our high layer should bring about the diffusion of light too. Among the optical density B of the layer at the zenith, the surface brightness of the sun $b_{\odot}$, and the sky b, the following relation holds (Link, 1948a)

$$b = 1.2 \times 10^{-5} B b_{\odot} . \tag{62}$$

At a time when the sun is approximately 5° below the horizon, the air layers at an altitude over 30 km are still illuminated, and the measurements yield

$$b = 10^{-9} b_{\odot} .$$

If, however, our layer of optical density $B = 0.005$ was to be found above this altitude, it should give the surface brightness

$$b = 6 \times 10^{-8} b_{\odot} ,$$

about 60 times larger than we see.

This implies that the high absorbing layer is thick and not all concentrated in higher altitudes. Its upper limit is revealed both by the measurements of the increase in the shadow and by the twilight phenomena. The curve of the twilight brightness of the sky as a function of solar depression displays a certain discontinuity in its course (Brunner, 1935; Link *et al.*, 1967) at altitudes of about 100 km, and this discontinuity might signify the upper limit of the layer.

XIX

METEOROLOGICAL ANALYSIS OF THE ECLIPSE

As early as in the last century, many observers pointed out various peculiarities in the structure of the internal parts of the shadow visible in mere telescopes without any further measuring instruments. It was, for instance, Kirch (1727) who, on the occasion of the eclipse of October 21, 1725, made an observation rather remarkable for that time. He observed in the internal part of the shadow a darker elliptical spot, flattened on the poles, and explained it (quite correctly) as due to a greater air density in the polar regions of our Earth. In many eclipses, a marked asymmetry in the density and color of the shadow was noted between the first and second part of the eclipse, in addition to considerable differences from one eclipse to another.

These phenomena in the internal parts of the shadow may, moreover, be explained by meteorological factors—such as the detailed structure of the atmosphere varying with seasons of the year, or the geographical latitude. As for the seasonal and geographical factors, the auxiliary shadows were computed for various conditions (Link and Guth, 1940; Bouška and Link, 1947); by means of the well-known aerological data (Ramanathan, 1929; Palmén, 1934); and from them the corresponding densities of the shadows were derived. The results indicate, as a matter of fact, certain systematic differences; but their values in most cases do not attain ± 0.1 in density. These computations were carried out for a cloudless Rayleigh's atmosphere.

Cloudiness may reveal itself either by the fact that the mean level of the family of cirrus clouds changes with the geographical latitude, or that in certain synoptical situations some parts of the effective terminator of the Earth's shadow may be more or less disturbed by clouds. Figure 32 represents schematically the isophotes of the auxiliary shadow in the month of November. The dashed line stands for the completely opaque isophote corresponding to the upper limit of cloudiness. The position of the solar disk S is chosen for the computation of the density of the shadow at the distance of 30′ and in the angle of position of 60°. The influence of the cloudy layer is rather evident. It will be greatest on the equator and smallest on the pole. The computation for the angles of position of 0° (pole), 40°S (intermediate latitudes), and 80° (tropics) gives an increase of
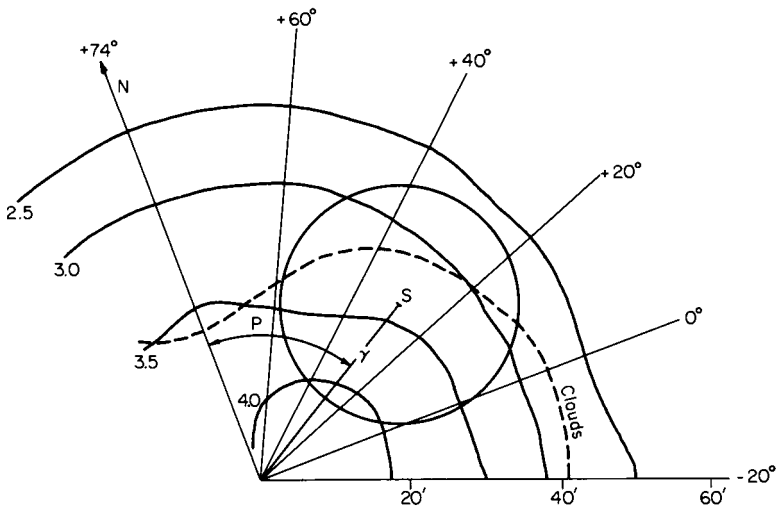
FIG. 32. Isophotes of the auxiliary shadow in November.

density due to cloudiness for +0.07, +0.16, and +1.05 at the distance of 25′ from the center of the shadow. A further analysis shows that, on the poles, cloudiness is ineffective as far as the distance of 35′ from the center of the shadow; and that to a distance of 20′, the interference of cloudiness is only weakly apparent. On the equator, however, cloudiness begins considerably to make itself felt from 35′ downwards.

This is a picture of the effect of average cloudiness. This is, of course, a maximum; since we have assumed an uninterrupted belt of clouds along the whole length of the effective terminator. In certain synoptic situations, essential deviations may occur—as becomes evident from the meteorological analysis of the eclipse of November 7, 1938. During this eclipse, a considerable flattening of the isophotes was found; and also density changes with time occurred in certain parts of the shadow (Guth and Link, 1939). These circumstances are well shown on Fig. 33.

For three epochs of Greenwich Mean Time, when the isophotes of the shadow were determined, Sekera used the Hamburg meteorological maps to establish the schematical profile of the three corresponding terminators of the Earth's shadow on the Earth (Fig. 34). It is evident from them that the parts of the terminator in lower latitudes were considerably disturbed by cloudiness, the effect being enhanced by a higher altitude of the clouds at these latitudes. On the other hand, in the polar regions of the terminator, the sky was almost cloudless. A further reduction of the shadow-brightness was related to the fact that the terminator swept into localities with increased cloudiness. This detailed meteorological
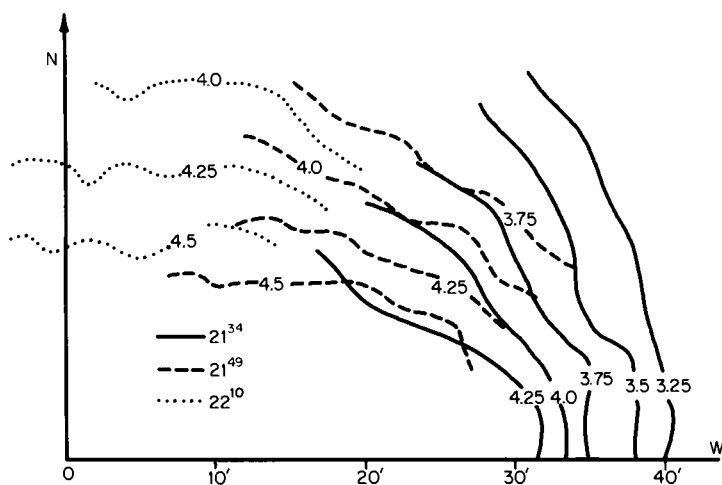
FIG. 33. Isophotes of the Earth's shadow during the eclipse on 7th November, 1938.

analysis of the lunar eclipse—the only one carried out so far—shows the way in which numerous anomalies in the central parts of the shadow might be explained. Such an analysis would frequently have to respect also various geographic factors —such as the screening caused by powerful mountain ranges projected in a
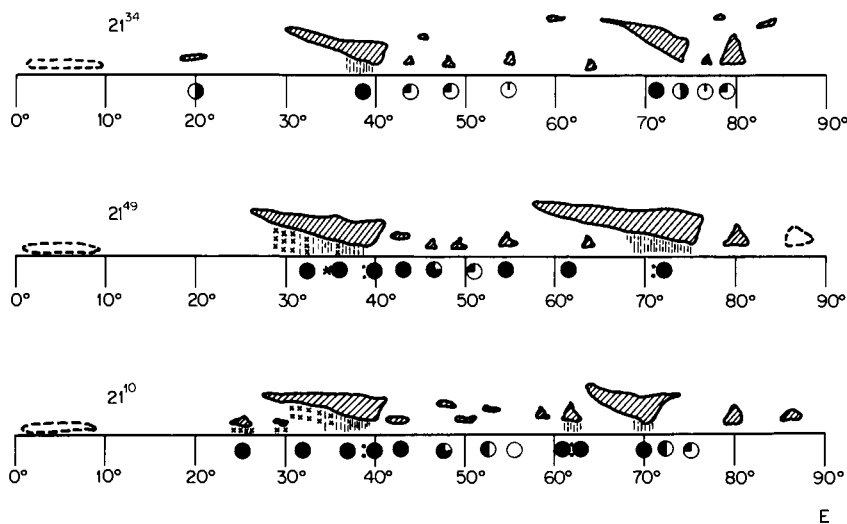
FIG. 34. Meteorological conditions on terminator of the shadow corresponding to the three moments in time as recorded on Fig. 33.

sufficient length into the terminator—as, for instance, the Cordilleras (Dufour, 1899).

In connection with the meteorological influences on the eclipse, a few additional words need to be said about the brightening of the shadow at its center which may be observed in numerous curves (Fig. 18). One of its causes may be the focusing of solar rays into the center of the auxiliary shadow which, though blurred, passes over into the actual shadow. Its effect is, of course, apparent only when the atmosphere is transparent up to 2 km (Section VII), and this is not always the case. A more appropriate explanation may be the fact that, in polar regions, refraction is always much greater than at lower latitudes; while the average altitude of the cloudy layer is also smaller. Thus, as soon as the polar parts of the terminator begin to exert an effect in the vicinity of the center of the shadow, more light than previously begins to penetrate into the center of the shadow and brings about its brightening. These polar "windows" offer a good explanation of the central brightening of the shadow (Link, 1948a; Švestka, 1949).

XX

INFLUENCE OF VOLCANIC ERUPTIONS

Apart from meteorological factors, the general atmospheric pollution of planetary extent by volcanic dust can produce a conspicuous increase in the density of the shadow, almost causing the disappearance of the eclipsed Moon. This idea, expressed on several occasions (Dufour, 1899; Flammarion, 1884), can be illustrated citing many dark eclipses in the past centuries (Link, 1961) and also the eclipses following the explosion of Mt. Agung in 1963 (Matsushima and Zink, 1964). Table XI records the principal cases of very dark eclipses.

In order to explain these phenomena we need to know if the disappearance of the Moon during the eclipse is theoretically possible at all. To have the illumination in the center of the shadow ($\gamma = 0$) reduce to zero, the illumination in the auxiliary shadow (Table III) must vanish up to the distance $r = R_{\odot} = 16'$ which, according to the parallax of the Moon, requires the opacity of the atmosphere up to the altitudes of 4 to 6 km, all over the terrestrial circumference of 40.000 km. The cloudiness can cause a local opacity of the atmosphere up to these altitudes along a limited part of the terminator, but not along its full length. The extension of cloudiness on a greater or smaller scale can, therefore, be expected to produce certain fluctuations of brightness, but not a complete disappearance.

Experience based upon various volcanic eruptions (Symons, 1888) has shown that, during great eruptions, volcanic dust will spread by the general atmospheric circulation around the whole Earth and also practically to all geographic latitudes. The altitude of the polluted layer attains at first some tens to fifty kilo-

TABLE XI

DARK LUNAR ECLIPSES[a]

No.	Date	Magnitude	Observers	Observations and volcanic explosions
1	1601 XII 9	11	Kepler	Eclipsed part invisible
2	1620 VI 15	18	Kepler	Moon invisible, stars visible
3	1620 XII 9	19	Cysat at Ingolstadt	Moon invisible
4	1642 IV 14	19	Many observers	Very dark or invisible. Explosion Avoe 1741 I, 1 km³ ashes
5	1761 V 18	18	Many observers	Very dark or invisible. Explosion Jorullo 1759 IX 28??
6	1816 VI 16	15	Lee at London, Eule at Dresden	Moon invisible. Explosion Tambora 1815, spring, 150 km³ ashes!!
7	1884 X 4	18	Many observers	Very dark and colorless. Explosion Krakatoa 1883 VIII, 18 km³ ashes
8	1902 X 16	18	Barnard and others	Very dark. Explosions 1902 V Mt. Pelée and St. Vincent, 1 km²; Santa Maria 1902 X 5 km³
9	1903 IV 11	12	Many observers	Very dark. Explosions as above
10	1913 III 22	19	Many observers	Very dark and colorless. Explosion
11	1913 IX 15	17	Many observers	Mt. Katmai 1912 VI-X, 21 km³
12	1963 XII 30	16	Many observers	Very dark and colorless. Explosion Mt. Agung 1963 III 17

[a] From Link (1961, 1963).

meters, followed by only a gradual downfall, which implies that the atmosphere does not clear up until after a number of months or a year. In this way, the condition for influencing the brightness of lunar eclipses by volcanic eruptions is indeed fulfilled. A sufficient absorption of the layer is yet another condition. Great volcanic eruptions produce from ten to a hundred cubic kilometers of ash and it may be presumed that at least 0.1 km^3 of it is the form of a fine ash whose particles (according to the size of Bishop's ring) have a mean radius a of approximately 10^{-4} cm.

We shall adopt as basis of our considerations the following model layer: altitude–20 km; range–half of the surface of the Earth (i.e., for instance between the latitudes 30°N to 30°S); the ejected mass–0.1 km^3 (i.e., 2×10^{14} gm). Now it is possible to compute the horizontal path $G'(h_0)$ and the total of the dust along this path, and the corresponding optical density D_V, which are listed in Table XII in comparison with the optical density D_R in Rayleigh's atmosphere (for green light).

If we consider that the surface brightness of the full Moon is 0.25 cd cm^{-2}, the same surface brightness subjected to the combined effect of pure atmosphere

TABLE XII

VOLCANIC DUST LAYER[a]

h_0 (km)	$G'(h_0)$ (km)	m (10^{-3} g)	D_V	D_R
0	1100	4.4	4.4	—
5	920	3.6	3.6	4.1
10	740	3.0	3.0	2.2
15	514	2.0	2.0	2.7

[a] From Link, 1963.

and volcanic dust can diminish to 10^{-8} of its value, which is less than the surface brightness of the night sky (10^{-8} cd cm^{-2}). To express it in a different way, a great volcanic eruption can bring about an effective disappearance of the eclipsed Moon.

The influence of the volcanic pollution of the atmosphere was actually measured on several occasions. After the eruptions relating to Krakatoa, 1883, St. Vincent, Mt. Pelée, Santa Maria, 1902–3, and after the Katmai eruption, 1912, the optical density at the zenith rose by 0.06 to 0.09, which corresponds to our model with an increase by 0.08.

Hansen and Matsushima (1966) discussed the eclipse of December 30, 1963 which, according to their measurements, was not only very dark but also colorless in central regions of the shadow. They computed several density curves for different atmospheric models with the following photometrical factors:

(a) *Rayleigh scattering*, giving the attenuation of the light and simultaneously an illumination by the diffused light in the atmosphere. This was done according to our general theory within the U.S. Standard atmosphere.

(b) *Ozone absorption* with two different total contents, 0.26 or 0.36 cm of O_3.

(c) *Aerosol scattering* in different tropospheric layers giving the attenuation and additional illumination according to formulas similar to (50).

In conclusion, a model of a homogeneous aerosol layer from the surface to 25 km with additional absorption of 0.084 m per air-mass was adopted, as it gives in the center of the shadow the great and approximately the gray density in agreement with observations. In other words, the main part of the illumination was there due to the diffusion in the aerosol layer.

It is very interesting to compare again this supposed aerosol extinction with its astronomical determinations based on Bouguer's method. Fortunately enough, Matsushima (1966b) found such measurements by Moreno and Stock (1964) and by Przybylski (1964), extended from March 1963 to the end of 1964.

After having treated two other eclipses which occurred during this time, namely the eclipse of June 25, 1964 measured by Bouška and Mayer (1965) and that of December 19, 1964 by Matsushima *et al.*, he obtained finally the following aerosol extinction coefficients:

Eclipse:	1963 XII 30	1964 VI 25	1964 XII 30
	0.08	0.06	0.044

Their place in the diagram of Bouguer values (Fig. 35) shows a quite satisfactory agreement between both methods. The great explosion of Mt. Agung
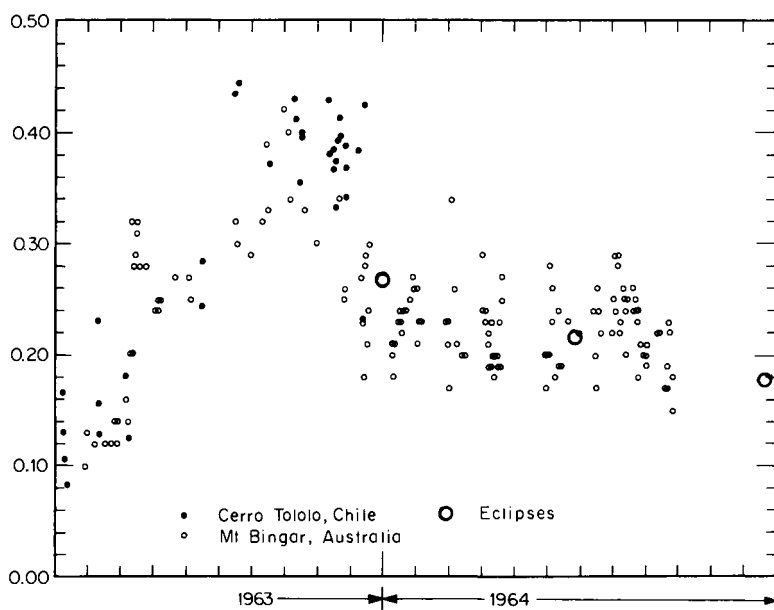
Fig. 35. Zenithal extinction obtained from Bouguer's graphs and from lunar eclipses (from Matsushima *et al.*, 1966).

(Bali) occurring on March 17, 1963 was without doubt the principal cause of this extended planetary atmospheric pollution.

In conclusion, many volcanic eruptions (Table XI) could be associated with a number of subsequent dark eclipses with the exception of the oldest phenomena where the reports of volcanic activity may be lacking. However, the Danjon relation may play some role for some of them, probably for the eclipse of May 18, 1761 and for the last eclipse, December 30, 1963.

XXI

LUNAR LUMINESCENCE

A great majority of terrestrial minerals, if irradiated by short-wave or corpuscular radiation, give off visible light by the process of luminescence. On the Moon, the conditions for luminescence—i.e., the presence of suitable minerals and of exciting radiation—are amply fulfilled. For this reason, it is only logical to look for the phenomena by which lunar luminescence might reveal its presence. The first way—even if to a certain degree indirect—offered itself in the lunar eclipses; and this was followed by the study of other phenomena explainable also by lunar luminescence.

In lunar eclipses, the possibility of luminescence revealed itself, first of all, in the penumbra (Link, 1946b, 1947b), when—in numerous cases—the computed density was found to be higher than the measured density. Let us compute the density of the penumbra from the viewpoint of the lunar observer who, in the penumbra, sees a partial solar eclipse caused by the Earth (Section XII). For the purposes of computing the illumination, we shall apply, in the first approximation, some results of the theory of eclipsing variable stars. If the ratio of the angular dimensions of the Sun and Earth observed from the Moon is

$$k = R_{\odot}'/\pi_{\leftmoon}, \tag{63}$$

and γ' denotes the angular distance of their centers, it follows that

$$\gamma' = \pi_{\leftmoon}\{1 + kp(k, \alpha; \kappa)\}, \tag{64}$$

where the "geometrical depth of the eclipse" $p(k, \alpha; \kappa)$ has been extensively tabulated by Tsesevich (1940) in series of k and α for several values of the coefficient of limb-darkening κ. The fractional loss of light α is, in turn, associated with the density of the penumbra by the relation

$$D = \log_{10}(1 - \alpha). \tag{65}$$

In the case of lunar eclipses, the value of k may range between 0.25 and 0.30, with a mean of 0.28. Furthermore, instead of the distance from the center γ', we find it advantageous to introduce the distance γ'' from the edge of the shadow to obtain the formula

$$\gamma'' = \pi_{\odot}\{109.05[1 + p(\alpha, 0.28; \kappa)] - 1\}, \tag{66}$$

which is practically independent of the lunar parallax.

From this formula, the Table XIII has been computed.

In these results, the effect of the ozone layer as well as that of the high-absorbing layer—both of which increase the density of the penumbra—have

TABLE XIII

DENSITIES OF THE PENUMBRA

γ''	D_1 Simplified theory	D_2 Complete theory	D_3 Preceding with ozone	D_4 Preceding with abs. layer	Observed	Observed—D_3 $= O - C$
0′		2.48	2.74	2.88	2.87	+0.13
1		2.18	2.31	2.40	2.10	—0.21
2		1.80	1.85	1.92	1.60	—0.25
3	1.35	1.41	1.44	1.47	1.34	—0.10
4	1.16	1.20	1.21	1.24	1.10	—0.11
5	1.02	1.04	1.05	1.08	0.96	—0.09

been neglected. Furthermore, the effect of the refraction image which decreases the density of the penumbra has also been neglected. Accurate computations have been carried out (Link, 1958) with new data concerning the upper atmosphere, obtained by rocket soundings (Rocket Panel, 1952), and with new data on the brightness of the solar disk in close vicinity of the edge (Heyden, 1954). The computations performed for the green light are to be found in Table XIII.

The influence of the neglected factors, influencing the density of the penumbra, becomes quite noticeable.

A total of 18 eclipses between 1921 and 1957 were investigated by Danjon's method (Section XV): 12 cases were found to exhibit an excess of light in the penumbra, 4 cases showed the reverse; and in two cases no adequate measurements in the penumbra were available (Link, 1958). In Table XIII, the median observed density of the penumbra is given. At a distance of 1′ 40″ from the edge of the shadow, for instance, the density of the shadow is by 0.27 less than the computed density—i.e., the illumination there is about twice as large as that given by the theory. In some individual eclipses these differences were even greater. Other measurements carried by photographic photometry in Italy (Cimino and Fortini, 1953; Cimino and Gianuzzi, 1955; Fortini, 1954, 1955) and by photoelectric photometry in Italy (Cimino and Fresa, 1958) and in Chile (Sanduleak and Stock, 1965) indicated a similar excess of light, so that its reality cannot be doubted.

In trying to explain this light excess, we shall proceed first by eliminating all known causes that might increase the illumination in the penumbra. One of them might be the scattering of light in the terrestrial atmosphere. Its value has already been computed (Section XIV); but now we wish to estimate it under very drastic conditions in order to estimate its upper limit. In order to obtain such a limit, let us assume that the terrestrial atmosphere up to an altitude of 120 km

radiates along the entire limb of the Earth with an intensity corresponding to a brightness 10^4 cd/m^2—which, on the Earth, would correspond to a very clear veiled sky in full sunshine. The parasitic illumination of the penumbra by such a ring should then be given by

$$e = \text{brightness} \times \text{solid angle} = 10^4 \times 3 \times 10^{-5} = 0.3 \quad \text{lux}, \tag{67}$$

while the direct illumination at the edge of the shadow is at least 100 lux. Hence, it follows that the scattering of light in the terrestrial atmosphere—even under the most favorable conditions—cannot increase appreciably the illumination in the penumbra.

In order fully to explain the light excess in the penumbra, we must, moreover, meet the objection (Barbier, 1959) that this phenomenon may be due to the special nature of the Moon's surface. As indicated by photometric measurements of the Moon (Fessenkov, 1962), each point of its disk attains the maximum brightness at the moment of opposition, and the brightness at each point in close vicinity to the maximum varies rapidly, rendering the maximum quite sharp. Thus, it might occur that an increase of brightness of the penumbra over the theoretical curve is caused precisely by the fact that our measurements are carried out at a time immediately preceding the opposition. Although such an explanation, similarly as in the case of light scattering in the atmosphere, is logically permissible, it is not quantitatively satisfying.

The opposition effect measured recently by Van Diggelen (1965) for several formations of lunar surface (Fig. 36) may be compared (Link, 1966) with the observed differences in the penumbra according to the last column of Table XIV. It is evident at first sight that the increase of brightness in the penumbra is much faster than in the normal light outside the eclipse. Therefore, the explanation
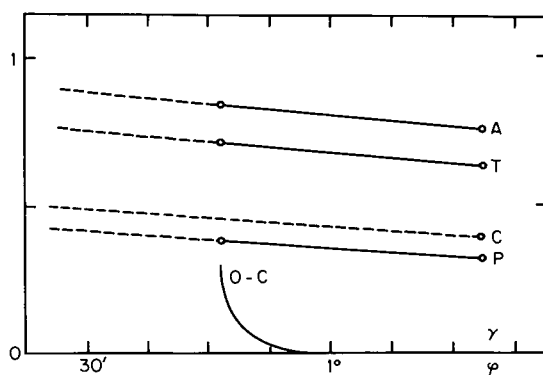
FIG. 36. Opposition effect on Aristarchus (A), Tycho (T), Copernicus (C), and Plato (P) compared with $O - C$ differences in the penumbra (from Link, 1966).

proposed by Barbier (1959) and repeated by many others does not hold, as has been shown from the old measurements of Mare Crisium by Wirtz (1915).

Thus when all known possibilities of an increase of the illumination in the penumbra have failed to account for the observed facts, let us resort to luminescence as a working hypothesis for the explanation of the observed excess. Its influence on the density of the penumbra becomes evident from the following consideration. Let us assume that 90 % of the light of the full Moon out of eclipse consists of sunlight scattered by the lunar surface, and that the residual 10 % arises from luminescence due to excitation of lunar minerals by short-wave or corpuscular radiation from the Sun. If the Moon in the penumbra were in such a position that at a certain point, 90 % of the photosphere becomes eclipsed, the first component decreases to one-tenth of its original value. The second component should, however, drop much less, since the source of exciting radiation—the corona—is less completely occulted by the Earth than the photosphere (Fig. 37). If, say, it will diminish to one-half of its original value, the
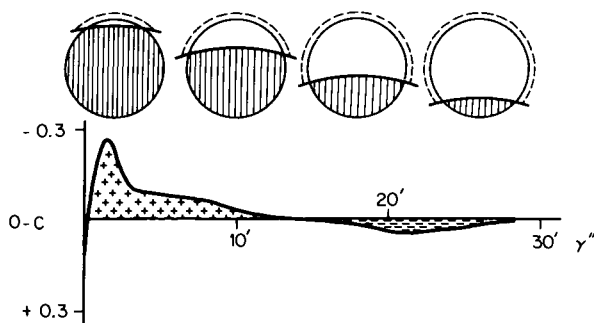
FIG. 37. Curve of $O - C$ differences in the penumbra as a function of the distance γ'' from the edge of the umbra. Above–the corresponding positions of the Earth and the Sun as seen from the Moon.

relative contribution of luminescence should rise to $5 : 14 = 35\%$ while, outside the penumbra, it amounted to 10 %. Towards the border of the shadow, these differences will only increase the relative importance of the luminescence. A similar consideration also holds good for corpuscular radiation; though in order to précis this argument, it would be necessary to investigate the form of the corpuscular shadow of the Earth with due regard to the Earth's magnetic field.

The assumption that at least part of the exciting radiation comes from the solar corona agrees with modern concepts of its structure, particularly its high temperature; in the light of which, the exciting radiation would be of the nature of soft X-rays. Their emission, probably, would not be distributed uniformly

over the solar disk, but a maximum may be expected at its edge, so that the source of radiation would be the outer ring surrounding the solar edge. A closer computation (cf. Fig. 37) shows that from the edge of the shadow, as far as a distance of about 16′ from its edge, the coronal ring is comparatively less eclipsed then the photosphere. At greater distances, this ratio is reversed. Therefore, it may be expected that in the internal part of the penumbra up to $\gamma' = 16'$, an excess of light will be found; and in the external parts for $\gamma' > 16'$, a light defect against the theory. The course of the mean values of the differences $O - C$ in Fig. 37 fulfills this expectation. In many cases, it was possible (cf. Dubois and Link, 1969) to detect the excitation of lunar luminescence by the K-3 plages when the limb of the Earth occulted these sources and the $O - C$ curve showed simultaneously the deviations from the normal form (see Fig. 8, p. 96, of Dubois and Link, 1969).

XXII

POSSIBLE MANIFESTATIONS OF LUNAR LUMINESCENCE OUTSIDE THE ECLIPSES

The fact of an excess of light in the penumbra lends support to an assumption that luminescence participates in the total radiation of the lunar disk at full Moon Since the cause of luminescence—the solar exciting radiation—seems to fluctuate in intensity with the solar activity, a corresponding fluctuation should be reflected also in luminescence, either in the penumbra, or also in the integrated light of the Moon. The number of photometrically measured eclipses being rather small, the fluctuations of the Moon's total light were investigated in the unobscured lunar disk (Link, 1947c).

Rougier's excellent series (1933) of photoelectric measurements was used as observational material. Rougier himself endeavoured to establish the dependence of the global brightness of the Moon on its phase. Each of his determinations of the Moon's brightness was based on an extrapolation beyond the terrestrial atmosphere by means of Bouguer's straight line, so that the effects of the fluctuation of the transparency of the terrestrial atmosphere were eliminated from the results. From all results, Rougier drew the mean curve of dependences of the stellar magnitude of the Moon on the phase angle. We used the deviations Δm of the individual measurements from this mean curve as an indicator of the luminescence, and searched for a correlation with the fluctuation of the solar constant Δc as determined by Abbot (1942a).

In the question of the fluctuations of the solar constant found by Abbot, there has been much discussion. The method used by Abbot for the determination of the solar constant was, in fact, theoretically quite correct. It is, however, rather

tedious and indirect, which exposed it to the danger of many systematic errors. The fluctuations of the order of 1 % reported by Abbot are comparatively small, and could not be traced in the reflected light of the planets where they also ought to appear. It must be pointed out, however, that the required accuracy lies at the limit of observational errors of planetary photometry affected as it is, by the intrinsic fluctuations caused, for instance, by variable cloudiness. The statistical criteria applied to Abbot's results render them at the limit of significance, but cannot exclude their reality.

However, when the correlation of the fluctuations of the solar constant with the fluctuations of the lunar light was investigated (Link, 1947c), positive results were obtained. The correlation coefficient (cf. Fig. 38) was found to be

$$r = -0.438 \pm 0.082 \quad \text{(m.e)}; \qquad n = 94,$$

indicating that the correlation is real. The negative sign of the correlation arises from the definition of the stellar magnitude and, from the physical point of view,
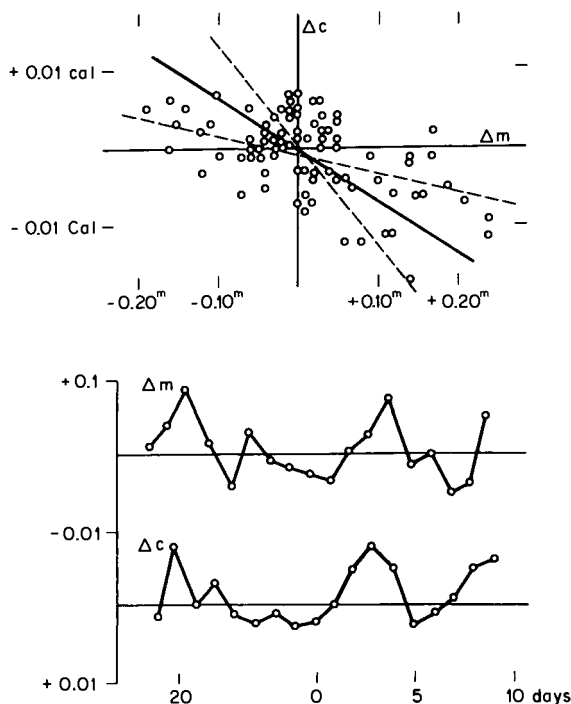
FIG. 38. Correlation of lunar luminescence Δm with the fluctuations of the solar constant Δc. Below: 27 days recurrence of values Δm and Δc.

must be taken as positive. Furthermore, from the mean inclination of the regression curves, it may be deduced that the fluctuation of 1% in the solar constant results in a fluctuation in the light of the Moon (i.e., of its luminescence component) of about 26%. The method is, in fact, much more sensitive than it was in the case of planets surrounded by atmospheres where one cannot expect to encounter any luminescence.

That small fluctuations of solar radiation in the visible part of the spectrum are accompanied by greater fluctuations in shorter wavelengths is, after all, a well-known fact. Thus Abbot (1942b) found that the ratio of the spectral variation to the solar constant increased towards shorter wavelengths. At the limit of its accessible spectrum (around 3500 Å), it attained the value 6 and apparently keeps on increasing, as indicated by Pettit's measurements (1933, 1938) and by rather numerous ionospheric results.

These results were, furthermore, treated so that the 27-day recurrence in the medians of the fluctuations of the solar constant Δc and of the brightness of the Moon Δm were investigated. The results reveal very neatly the mutual relation (Fig. 38)—with one peculiarity, however: namely, that three maxima in the curve Δm lag one day behind the corresponding maxima of Δc. This seems to indicate that the excitation of the lunar luminescence is caused by corpuscular solar radiation, reaching the Earth from the Sun with a time lapse of the order of 1 to 2 days (Link, 1951).

A further possibility for the study of luminescence offers itself in the method of the depth of Fraunhofer's lines which was proposed by Link (1950b) at the Congress of I.A.U. in Zürich, 1948. If i stands for the central intensity of the line and I for the intensity of the continuous background, then (cf. Fig. 39) the ratio

$$R = i/I \tag{68}$$
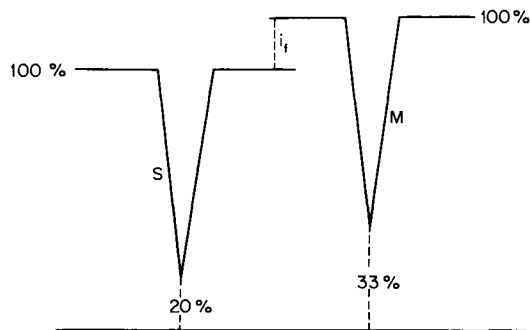
Fig. 39. Central intensity of a line in the solar (S) and lunar (L) spectrum in the presence of luminescence.

defines the so-called depth of the line, for instance, in the spectrum of the Sun. The same line in the spectrum of the Moon will, however, possess the depth

$$R' = (pi + i_f)/(pI + i_f), \tag{69}$$

where p is the reflection coefficient of the lunar region at the particular wavelength, and i_f the intensity of the luminescence component which we express by ratio

$$\rho = i_f/pI, \tag{70}$$

or also

$$\rho = (R' - R)/(1 - R'). \tag{71}$$

The depth of one and the same line can be measured with the same instrument in the spectra of the Sun and Moon; and with the help of Eq. (71), we can compute the relative participation ρ of the luminescence component in the light of the Moon.

These kinds of measurements were carried out by Dubois (1956, 1957, 1959) and Kozyrev (1956). They revealed a distinct luminescence in some parts of the Moon, confined to certain emission bands of the spectrum. Out of 86 spots investigated by Dubois, luminescence was reported in 47 cases. It is interesting to note that Kozyrev (1956) found by the same method a comparatively intense luminescence in the region of the K and H lines in the light of the rays surrounding the crater Aristarchus; a fuller description of his observations can be found elsewhere (Kozyrev, 1962).

XXIII

THE VARIATION IN THE BRIGHTNESS OF THE ECLIPSE

The observed historical eclipses of the Moon have indicated that the brightness of such phenomena on record varied within very wide limits. Thus there are reports of eclipses during which the Moon disappeared entirely from sight, or was only seen with great difficulty. This, for instance, appeared to be the case for the eclipse on May 5, 1110, observed in England. The eclipse on June 15, 1620, was, according to Kepler, very dark and the visibility of the Moon was greatly reduced. In the course of the eclipse of April 14, 1642, the Moon disappeared completely from Hevelius's view; just as it did from Wargentin's during the eclipse on May 18, 1761; and from Lee's on June 10, 1816. On the other hand, there are also records of very bright eclipses characterized by rather lively colors. This was apparently true of the eclipse of September 9, 1718, as observed by Cassini; and of that of March 18, 1848, observed at a great number of observatories.

Similar reports on the fluctuations of the brightness of the eclipses prompted Danjon (1920a) to a treatment of the dependence of the brightness and color of eclipses on the solar activity. In order to be able numerically to express verbal descriptions of eclipses, he introduced a luminosity scale of eclipses ranging from $L = 0$, when the Moon is almost or wholly invisible, to $L = 4$, when the eclipsed Moon is exceptionally bright. Together with the brightening of the Moon, its color varied from a neutral black $L = 0$, to brown, brick-red, red, up to orange at $L = 4$. Danjon applied his procedure first to seventy eclipses from the period of 1823 to 1920. The results showed an interesting dependence of Danjon's degree of brightness L on the phase ϕ of the solar activity, computed in fractions of the eleven-year cycle from the minimum (Fig. 40). This dependence was
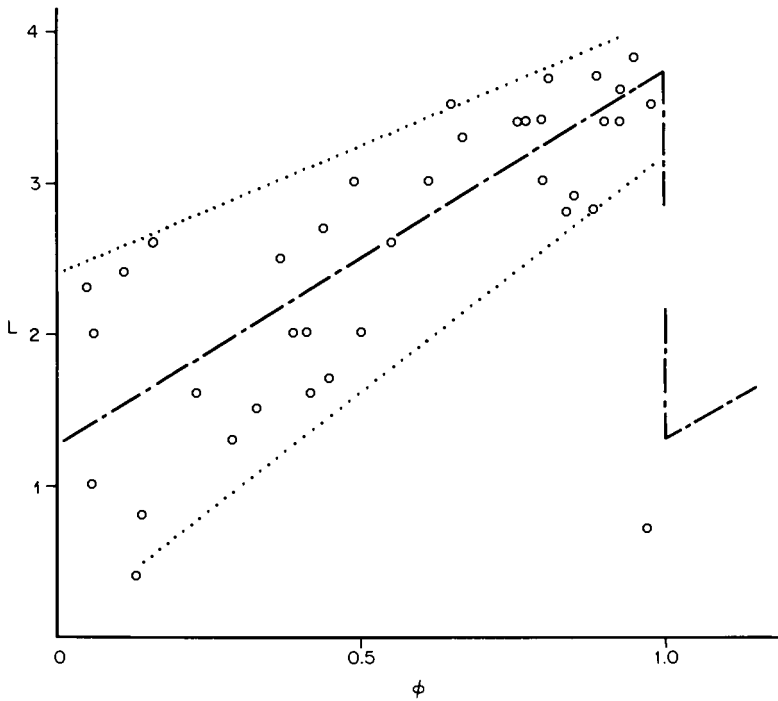
FIG. 40. Danjon's relation of the dependence of the luminosity of lunar eclipses L on the phase ϕ within the 11 year solar cycle.

supplemented and confirmed by de Vaucouleurs (1944) on the basis of the results obtained from observations of forty-seven eclipses between 1893 and 1943. At the minimum of solar activity, the dependence exhibits a distinct discontinuity. The eclipses preceding this minimum are brightest, and those following it rapidly

become fainter. From one minimum to another, the brightness of the eclipse increases and diminishes quite linearly, giving rise to a saw-toothed curve.

Later, Danjon (1920b) extended these investigations back to 1583, and reversed the whole procedure: Namely, he took the dependence given by the curve in Fig. 40 as proved, and used its jump at the minimum for the determination of the epochs of the minima of solar activity in the past. The minima thus determined may be expressed by the equation

$$\min = 1584.8 + 10.87E + 1.7 \sin 2\pi[(t - 1608)/136], \qquad (72)$$

which agrees well with the lunar minima of light and equally well with the actual minima of solar activity according to Wolf, as long as we confine ourselves to periods following 1823. At earlier epochs, Danjon–Wolf's deviations are much greater, which might be due to the bad quality of solar observations. In fact, we must point out with Danjon, that the sunspot numbers before this year were obtained not by routine observations inaugurated by Schwabe in 1823, but from compilation of fragmentary observations collected by Wolf in old records. For example, the discovery of sunspots in 1611 was highly improbable as it is the year of solar minimum, according to Wolf. After 1823, the agreement of eclipse brightness and solar minima becomes suddenly very good, especially, when we add according to Bell and Wolbach (1965) a correction of +0.4 year to the initial epoch of Eq. (72). These authors have extrapolated the corrected equation into the future in order to obtain the next solar minima in 1975; 1986; 1997; 2008 and 2020.

One might be intrigued by the average length of the solar cycle, 10.87 years, in contrast with generally adopted value of 11.1 years. We must first object that during the best observed interval, 1823–1954 the average value of sunspot cycles is only 10.9. Second independent research on solar activity in the past by Link (1955a) and by Brooks (1928) on the basis of climatic variations, led to the average value of 10.85 years with fluctuations in limits between 12.1 and 9.6 years. Thus one of Maunder's objections (1921) is unsubstantiated as well as the others against the essence of the method itself, as has been shown by Danjon (1925).

Fisher (1924) too raised objections against Danjon's results. He treated all eclipses observed within the period 1860–1922 on a different scale of brightness which was based on the simple visibility of lunar details by telescopes of increasing aperture. This criterion allows, however, alternative interpretations; since the visibility of a diffuse object depends, among others, also on the magnification of the telescope. Nevertheless, some of Fisher's results must be regarded as a positive contribution—like, for instance, his proof of the dependence of the brightness of the eclipse on the season of the year (the hibernal eclipses being the brightest)—as also confirmed by Smith's findings (1885)—or the effect of volcanic eruptions (1883, 1902, 1912) on the brightness of the eclipse. Fisher confirmed

neither Danjon's relation, nor did he find the dependence of total eclipses on the parallax, which is in direct conflict with both theories (see, e.g. Table V) and observation (Švestka, 1950). On the whole, an analysis of this controversy (Link, 1956c) allows us to state that Danjon's results are of a great value in the field of the study of lunar eclipses and should become the starting point of many new investigations.

It would, however, be very desirable to verify Danjon's results both by a new
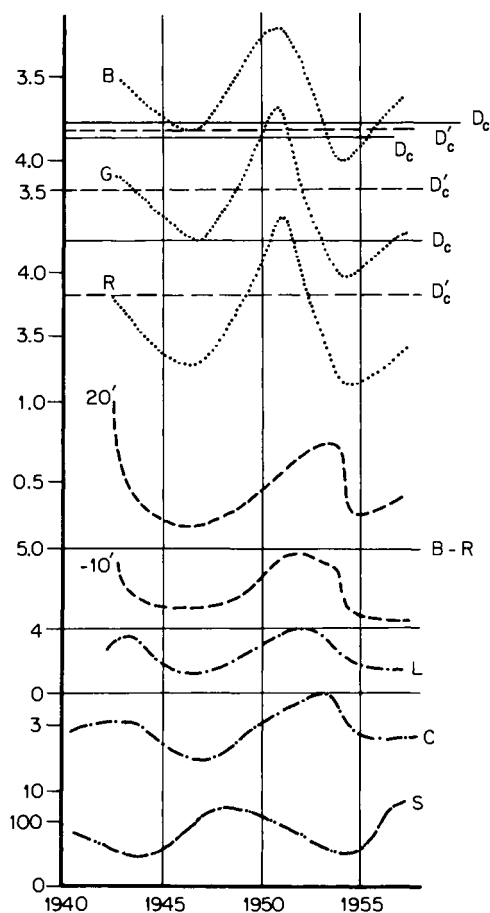
FIG. 41. Changes in brightness of lunar eclipse according to measurements of the shadow density. Above: changes of the shadow density at the distance $\gamma'' = -10'$ for the blue (B), green (G), and red (R) color. D_c density computed for Rayleigh's atmosphere, D_c' same but with ozone. Center: differences B–R for $\gamma = 20'$ and for $\gamma'' = -10'$. Below: luminosity curve L, curve of the brightness of the earthlight on the Moon C, and curve S of sunspots.

treatment of the historical material, extended if possible to further sources, as well as by its comparison with the results of the measurement of the shadow densities.

For the time being, only the second part of this program, namely, the comparison with measurements of the optical density, has been carried out (Link, 1960a). This has been performed for eighteen eclipses measured by Danjon's method from 1921 on. The results are given in Fig. 39 for three spectral regions, the blue (B), green (G), and red (R). The results confirm, on the whole, Danjon's law—particularly as far as the coloring of the shadow is concerned—characterized numerically by the difference B − R. In the limb regions ($\gamma'' = -10'$) as well as closer to the center ($\gamma = 20'$) in 1944 and 1954, an abrupt decrease of the differences B − R could be noted in the vicinity of the solar minima, when, in fact, the predominance of the red color in the shadow steeply diminished. Both curves run, on the whole, synchronously with the curve of the brightness of the eclipse L on Danjon's scale. As for the density of the shadow, a similar fluctuation appears in the limb regions ($\gamma'' = -10'$), with the phase shifted with regard to the color curve B–R. The bright eclipses in 1950 do not stand out on the color curve and the again marked maximum of color around 1953 coincides with low brightnesses of the shadow.

A part of the density curves lies above the theoretical limit of density for Rayleigh's atmosphere with ozone D_c' (Fig. 41) and sometimes even above the density D_c only in Rayleigh's atmosphere. Such a phenomenon may be ascribed only to the additive light in the shadow. The value of this additive light in the case of a very bright eclipse on the September 26, 1950, is shown in Table XIV.

As long as we wish to explain this excess by atmospheric light diffusion alone, it becomes apparent that this diffusion may yield only a very small part—a maximum of 10×10^{-6}. In this case, we encounter again a similar phenomenon, as in the case of the penumbra, which likewise finds a logical explanation in the

TABLE XIV

Eclipse of 26.9.1950 at $\gamma = 20'$

	Blue	Green	Red
Density in Rayleigh's atmosphere	4.60	3.56	3.02
Density due to the ozone	0.03	0.33	0.43
Computed density	4.63	3.89	3.45
Observed density	3.68	3.40	3.09
Obs.–Comp.	—0.95	—0.49	—0.36
Observed illumination · 10^6	209	398	813
Computed illumination · 10^6	23	129	355
Obs.–Comp. · 10^6	+186	+269	+458

luminescence either directly excited in the shadow (most probably by corpuscular radiation) or in the form of an afterglow of the luminescence from the period preceding the eclipse.

A recent research by Matsushima (1966c) on the eclipse brightness in association with the geomagnetic activity has shown their correlation (Fig. 42). There
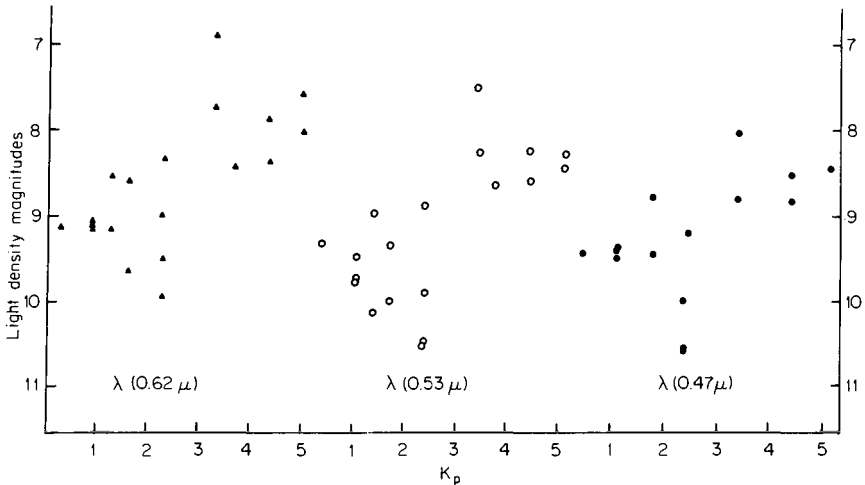
FIG. 42. Correlations of the eclipse brightness in the umbra at 10′ from its edge with the planetary geomagnetic index k_p (from Matsushima, 1966c).

is plotted the shadow density at 10′ from the edge as compiled by Link (1960a) in dependence on the planetary geomagnetic index K_p. As this index is an image of the corpuscular flux obtaining in the neighborhood of the Earth, it may be also taken in the analogous manner for the eclipsed Moon. This is in fact the case of Fig. 42. Although the scatter is not negligible, we can see there a definite decrease of the shadow density with increasing K_p. This may be interpreted as the luminescence excited by the corpuscular bombardment of the lunar surface. Further evidence of this bombardment is provided by the annual variation of the light excess in the umbra (cf. Dubois and Link, 1970). The observed curve shows two maxima at equinoxes and minima at solstices like the curve of auroral frequency or of geomagnetic activity, which are beyond doubt related with the corpuscular incidence near the ecliptic where the eclipses take place (see Fig. 6, p. 470, of Dubois and Link, 1970).

In further discussion, Matsushima (1966a) considers the energy necessary for this process. Taking the relation of the velocity of the solar plasma derived by Snyder *et al.* (1963) during the Mariner II flight with K_p, we can calculate the density of the proton flux necessary to produce the additional illumination in the umbra by the luminescence. Its value is by an order of magnitude greater

than the experimental value. The same difficulty arises when we discuss the variation of the lunar brightness outside the eclipses. Photometric observations of the Moon by Gehrels *et al.* (1964) indicate an increase of 36 to 27% during November 17–18, 1956 when the index K_p was as high as 2^+ and 5^- and the normal light November 19, with $K_p = 0^+$. The deficit in energy balance is here at least three orders of magnitude. The same difficulty arises when we examine lunar samples under artificial excitation in terrestrial laboratories.

These and similar considerations lie, however, beyond the scope of the present chapter and their discussion is to be found elsewhere (Sidran, 1968).

Finally, let us mention an interesting relation between the changes in brightness of the Earth-light with the solar activity, and also with the brightness of lunar eclipses, as pointed out by Dubois (1958). The curve of the Earth-light designated by C in Fig. 41 shows the same features as the curve B–R of the color or the curve L of the brightness in Danjon's scale. Two interpretations of these facts are possible, according to the following schemes:

Brightness of lunar eclipses	Brightness of the Earth-light	Heliogr. latitude of sunspots	Lunar luminescence	Aerosols in the atmosphere
maximum	maximum	low	maximum	maximum
minimum	minimum	high	minimum	minimum

The first interpretation (Link, 1947b) seeks the explanation of lunar variations in the luminescence of the Moon's surface excited by corpuscular emission of the Sun. The second interpretation (Vassy, 1958) adopts the hypothesis of changing percentage of aerosols in the terrestrial atmosphere with solar activity. The light diffused by it causes both the lunar fluctuation, either increasing the illumination during the eclipses or increasing the terrestrial albedo outside the eclipses. The common point of both interpretations is the abrupt change of sunspot latitude which arrives, according to Spörer's law, at the solar minima. Before the solar minimum, when the centers of solar activity (sunspots) and simultaneously the sources of corpuscular streams are in lower heliographic latitudes, this radiation may reach the Earth or the Moon. After the minimum has been attained, the centers suddenly rise into higher latitudes, and the corpuscular flux weakens abruptly in the terrestrial space.

XXIV

FURTHER PHENOMENA DURING LUNAR ECLIPSES

Lunar eclipses cause an abrupt discontinuity in the comparatively calm state of the lunar body, not too disturbed by the slow alternation of the days of fourteen

terrestrial days in length, and of the equally long nights. By the grace and accident of the phenomena of eclipses, other observations can be made which would otherwise be impossible. In what follows we shall briefly mention some of these possibilities in turn.

A. *Temperature of the Lunar Surface*

The temperature undergoes large changes during eclipses, since its source—the solar radiation—is suddenly cut off for several hours. Measurements of the long-wave radiation may provide us with certain information on the temperature-drop and simultaneously also on the thermal characteristics of the surface layers.

B. *Changes of Brightness*

Changes of brightness of some of the details of the lunar surface have been reported already from time to time, but their reality could so far not be confirmed Relatively greatest attention attached to the crater Linné, by its changes both during eclipses as well as when the lunar disk was not obscured. During eclipses, its diameter seems to increase, but it cannot be decided whether this is an actual or only an apparent change caused by physiological factors. Link (1947a) pointed out the possible relevance of luminescence in this connection.

C. *Dust-Shadow*

During the eclipse of the November 16, 1910, a dust-shadow was observed by a group of students under the guidance of Babitchev at the Odessa Observatory. From the obscured Moon standing some 40° above the horizon, a shadowy cone issued which could be traced almost to the horizon. Witkowski, also present among the observers, later explained this phenomenon (1949) as caused by a cloud of cosmic dust in which the lunar body created a dust-free space—something resembling a shadow. This explanation appears to be very probable, if we recall that it is around November 16 that the maximum of Leonids takes place, and that during the eclipse of that date (Link, 1956d), an extraordinary great shadow-enlargement was found attaining 3.7% (Section XVIII). In many cases, some extension of the terrestrial shadow beyond the lunar disk could be observed.

D. *Observations of Faint Bodies*

Observations can be made in the vicinity of the eclipsed Moon—as, for instance, the occultations of stars too faint to be seen under normal circumstances. The search for natural or artificial satellites of the Moon would obviously be facilitated at such times.

XXV

ALLIED PHENOMENA

An analysis of the contemporary problems in astronomy shows (Barbier, 1959) that their classification necessarily must be twodimensional, partly depending on the subject (for instance, in rows), and partly in the working method (in columns)—be this method experimental or theoretical. These circumstances may be illustrated also in our case of lunar eclipses, when our methods embrace other subjects also. Thus, if we proceed in the columns of the method we arrive at the following applications.

A. *Optical Occultations of Stars by the Moon and Planets*

We proceed from the simplified equation for the attenuation of the light by refraction [(Eq. (22)], as has been shown previously (Section VII).

B. *Radio Occultations by the Moon and Planets*

In the radiowave region, again Eq. (17) and the equation for the refraction (26) —after due conversions to the required form—may be used, provided that the Moon has an ionosphere. According to a theory formulated by Link (1955b, 1956a), minute traces of a lunar atmosphere are discernible. The method is about a thousand times more sensitive than the optical method used by Lyot and Dollfus. This method was used by Elsmore(1955), and he found the upper limit 10^{-12} of the density of terrestrial atmosphere. Mariner missions toward Venus and Mars belong to the same category of exploration.

C. *Radio Occultations by the Solar Corona*

Machin and Smith (1951) suggested using the occultation of the radio source in Taurus by the solar corona for the determination of the coronal structure. Link (1952) developed a theory of occultation, based on lunar eclipses, that later served for the interpretation of the observed phenomenon (Hewish, 1955).

D. *Transits of Venus over the Sun*

These phenomena are an analogy of the lunar and solar eclipses. Their observations, based on our theory, are capable of throwing a new, interesting light on some problems concerning the atmosphere of Venus (Section XII).

E. *Eclipses of Other Satellites*

To these phenomena, our method may be very conveniently applied (Link, 1936a). Its application to photometric measurements (Eropkin, 1931) suggests the presence of a very high-absorbing layer that lies beyond Jupiter's atmosphere and is reminiscent of some rudimentary remnant of a ring. Similar absorption but within the atmospheric limit seems to exist in the Martian atmosphere according to Rakos (1966).

However, the eclipses of Earth's artificial satellites are now taking over (Link, 1962) the heritage of lunar eclipses. The outlook is very promising for subsequent study of both the theoretical and observational sides of the problem.

F. *Einstein's Light Deflection*

The deflection of light in the gravitational field of a sufficiently large mass entails consequences similar to those of the refraction of light in the atmosphere. A more detailed theory (Link, 1936b, 1937; Zwicky, 1937) leads again to the prediction of certain observable phenomena. It seems that the peculiar object in Serpens—observed by Hoag (1949) as a perfect halo 17″ in radius which surrounds a diffuse central image—may possibly belong to this category.

XXVI

CONCLUSIONS

At the end of our survey of the phenomena associated with lunar eclipses, let us summarize the most important scientific notions, arrived at with the aid of these phenomena, or independently confirmed by them.

1. Spherical shape of the Earth (4th century B.C.).
2. Relative dimensions of the system Earth–Moon–Sun (3rd to 2nd centuries B.C.).
3. Theory of the lunar motion (chiefly from the 2nd century A.D. onwards up to the present).
4. Determination of the geographical longitudes (16th to 18th centuries).
5. Chronology.
6. Enlargement of the Earth's shadow (beginning with the 17th century; particularly, however, at the end of the 19th century and at present).
7. Distribution of ozone in the atmosphere.
8. Existence of a high-absorbing layer and its relation to interplanetary matter.
9. Light excess in the shadow showing an existence of lunar luminescence.
10. Thermal characteristics of the lunar surface crust.

11. Variations of the brightness of lunar eclipses with the solar activity.
12. Application of the theory of lunar eclipses to optical and radio phenomena in the solar system (atmospheres of the planets and of the Moon, solar corona) and to artificial satellites.

Points 7 to 12 represent achievements of the last four decades.

The perspectives of the further development of our scientific discipline lie apparently in a continuation of the tactics hitherto followed and expressed by the motto: "To proceed from the known to the unknown." Today, when we may study the upper atmosphere directly, thanks to the rocket technique, it will be possible to compute on the basis of the data thus obtained the average structure of the Earth's shadow with a comparatively great accuracy. A comparison of these data with the results of detailed measurements will provide us with a deeper knowledge of the atmosphere. An advantage of this procedure consists in the fact that lunar eclipses give us a simultaneous, geographically extended picture of the terrestrial atmosphere, while rocket soundings are, and always will be, restricted to specific localities.

It would be erroneous to conclude that the development of the technique of artificial satellites will lead to a complete and direct understanding of all problems of the whole terrestrial atmosphere. There remains, in fact, a zone between 40 and 200 km of altitude, where the atmosphere is too dense to allow a protracted stay of these space laboratories flying at high velocities in accordance with the laws of celestial mechanics; but where the atmosphere is again too tenuous to support floating balloons. This zone (40–200 km) is, however, of a great importance from scientific as well as technical points of view, and research about it remains, for the time being, almost exclusively the domain of high-altitude rockets together with the classical indirect methods used so far (including that of the lunar eclipses) which have the best opportunity to supplement and to support each other.

References

Abbot, C. G. (1942a). *Ann. Astrophys.* **6**, 164.
Abbot, C. G. (1942b). *Ann. Astrophys.* **6**, 85.
Bao-Lin, L. (1964). *Acta Astron. Sinica* **12**, No. 1.
Barbier, D. (1959). Private communication.
Bauer, E.,and Danjon, A. (1923). *Bull. Soc. Astron. France* **37**, 241.
Beer, W., and Maedler, J. H. (1834). *Astron. Nachr.* **11**, 290.
Bell, B., and Wolbach, J. G. (1965). *Icarus* **4**, 209.
Bouška, J. (1948). *Bull. Astron. Inst. Czech.* **1**, 37 and 75; **2**, 28; **7**, 85; **9**, 245; **11**, 145; **17**, 92; **21**, 61.
Bouška, J., and Link, F. (1947). *Compt. Rend.* **224**, 1483.
Bouška, J., and Mayer, P. (1965). *Bull. Astron. Inst. Czech.* **16**, 25.
Bouška, J., and Švestka, Z. (1950). *Bull. Astron. Inst. Czech.* **2**, 6.

Bowen, E. G. (1953). *Australian J. Phys.* **6**, 490.
Bowen, E. G. (1956). *J. Meteorol.* **13**, 142.
Bowen, E. G. (1957). *Australian J. Phys.* **10**, 412.
Brooks, C. E. P. (1928). *Mem. Roy. Metereol. Soc.* **2**, No. 12.
Brosinsky, A. (1888). "Über die Vergrösserung," etc. Göttingen.
Brunner, W. (1935). *Publ. Sternw. Zurich* **6**, 81.
Buchanan, R. (1907). "The Mathematical Theory of Eclipses," p. 202. Philadelphia, Pennsylvania.
Cabannes, J. (1929). "Sur la diffusion de la lumière," p. 167. Paris.
Cabannes, J., and Dufay, J. (1927). *J. Phys. Theoret. Appl.* **8**, 125.
Calvisius, S. (1650). "Opus Chronologicum." Francoforti ad Moe.
Cassini, J. D. (1740). "Tables astronomiques," p. 34. Paris.
Chambers, G. F. (1899). "The Story of Eclipses." London.
Chauvenet, W. (1908). "Manual of Spherical and Practical Astronomy," p. 542. Philadelphia, Pennsylvania.
Cimino, M., and Fortini, T. (1953). *Atti Accad. Nazl. Lincei* **14**, 619.
Cimino, M., and Fresa, A. (1958). *Atti Accad. Nazl. Lincei* **25**, 58.
Cimino, M., and Gianuzzi, M. A. (1955). *Atti Accad. Nazl. Lincei* **18**, 173.
Danjon, A. (1920a). *Compt. Rend.* **171**, 127.
Danjon, A. (1920b). *Compt. Rend.* **171**, 1207.
Danjon, A. (1921). *Astronomie* **39**, 272.
Danjon, A. (1925). *Bull. Soc. Astron. France* **39**, 272.
Danjon, A. (1928). *Ann. Obs. Strasbourg* **2**, 1.
de Vaucouleurs, G. (1944). *Compt. Rend.* **218**, 655 and 805.
Dobson, G. M. B. (1930). *Proc. Roy. Soc.* **A 129**, 411.
Dollfus, A. (1962). *In* "Physics and Astronomy of the Moon" (Z. Kopal, ed.), 1st ed., Chap. 5. Academic Press, New York.
Doublet, E. (1922). "Histoire de l'astronomie," p. 279. Paris.
Dubois, J. (1956). *Astronomie* **70**, 225.
Dubois, J. (1957). *J. Phys. Radium* **18**, Suppl., 13.
Dubois, J. (1958). *P.V. Soc. Sci. Phys. Nat. Bordeaux* **24**, IV.
Dubois, J. (1959). *Rozpr. Cesk. Akad. Ved.* **69**, No. 6.
Dubois, J., and Link, F. (1969). *The Moon* **1**, 85.
Dubois, J., and Link, F. (1970). *The Moon* **1**, 462.
Dufour, C. (1899). *Astronomie* **13**, 115.
Du Séjour, (1786). "Traité analytique des mouvements apparents des corps célestes." Paris.
Elsmore, B. (1955). "I.A.U. Symposium on Radioastronomy," No. 4, p. 403.
Eropkin, D. J. (1931). *Z. Astrophys.* **3**, 163.
Fabry, C. (1929). *J. Observateurs* **12**, 1.
Fessenkov, V. G. (1937). *Astron. J. (Moscow)* **14**, No. 5.
Fessenkov, V. G. (1962). "Physics and Astronomy of the Moon" (Z. Kopal, ed.), 1st ed., Chap. 4. Academic Press, New York.
Fisher, W. (1924). *Smithsonian. Inst. Misc. Collections* **76**, No. 9.
Flammarion, C. (1884). *Astronomie* **3**, 401.
Fortini, T. (1954). *Atti Accad. Nazl. Lincei* **17**, 209.
Fortini, T. (1955). *Atti Accad. Nazl. Lincei* **18**, 65.
Frisch, C. (1858). "J. Kepleri Opera Omnia," Vol. 2, p. 297. Francoforti.
Gaubil, A. (1740). "Observations Astronomiques," Vol. 3. Paris.
Gehrels, T., Coffeen, T., and Owings, D. (1964). *Astron. J.* **69**, 850.

Götz, P. (1931). *Beitr.Geophys.* **31**, 119.
Götz, P. (1940). *Meteorol. Z.* **57**, 415.
Green, A. (1964). *Appl. Opt.* **3**, 203.
Greenstein, J. L. (1937). *Harvard Circ.* No. 422.
Guth, V., and Link, F. (1939). *Z. Astrophys.* **18**, 207.
Hansa, M., and Zacharov, I. (1958). *Bull. Astron. Inst. Czech.* **9**, 236.
Hansen, J. E., and Matsushima, S. (1966). *J. Geophys. Res.* **71**, 1073.
Hartmann, J. (1891). *Abh. Saechs. Ges. (Akad.) Wiss.* **17**, 365.
Hausdorff, F. (1895). *Ber. Verhandl. Saechs. Akad. Wiss. Leipzig, Math.-Naturw. Kl.* **47**, 401.
Heath, T. (1913). "Aristarchos of Samos." Oxford Univ. Press, London and New York.
Hepperger, J. (1895). *Sitzber. Akad. Wiss. Wien, Math.-Physik. Kl. Abt.* **54**/*II*.
Hewish, A. (1955). *Proc. Roy. Soc.* **A228**, 238.
Heyden, F. J. (1954). *Astrophys. J.* **118**, 412.
Hoag, A. (1949). *Astron. J.* **55**, 170.
Humphreys. G. W. (1929). "Physics of the Air." New York.
Kirch, A. (1727). *Misc. Berol.* p. 228.
Koebke, F. (1951). *Bull. Soc. Amis. Sci. Poznan* **11**.
Koomen, M. J., Lock, C., Packker, D. M., Scolnik, R., Tousey, R., and Hulburt, E. O. (1952). *J. Opt. Soc. Am.* **42**, 353.
Kosik, S. M. (1940). *Bull. Tashkent Astron. Obs.* **2**, 79.
Kozyrev, N. A. (1956). *Izv. Crim. Astrophys. Obs.* **16**, 148.
Kühl, A. (1928). *Physik. Z.* **29**, 1.
Lahire, P. (1707). "Tabulae astronomicae." Paris.
Lalande, J. F. (1783). *Astronomie* **2**, 337.
Lambert, F. (1782). "Briefwechsel." Berlin.
Langrenus, M. F. (1644). "Tractatus de vera longitudine," etc. Antverpiae.
Laplace, S. P. (1860). "Oeuvres complètes," Vol. 6, Part I, p. 16. Paris.
Legentil, G. J. H. (1755). *Mem. Acad. Sci. (Paris)* p. 36.
Lemonnier, P. C. (1746). "Institutions astronomiques," p. 251. Paris.
Link, F. (1929). *Bull. Obs. Lyon* **11**, 229.
Link, F. (1933). *Bull. Astron. Paris* **8**, 77.
Link, F. (1936a). *Bull. Astron. (Paris)* **9**, 227.
Link, F. (1936b). *Compt. Rend.* **202**, 917.
Link, F. (1937). *Bull. Astron. (Paris)* **10**, 75.
Link, F. (1943). *Beitr. Geophys.* **60**, 139.
Link, F. (1946a). *Ann. Astrophys.* **9**, 227.
Link, F. (1946b). *Compt. Rend.* **223**, 976.
Link, F. (1947a). *Bull. Astron. (Paris)* **13**, 1.
Link, F. (1947b). *J. Soc. Math. Phys. Tcheque* **72**, 65.
Link, F. (1947c). *Colloq. Lyon, C.N.R.S. Paris* **1**, 308.
Link, F. (1948a). *Ann. Geophys.* **4**, 47 and 211.
Link, F. (1948b). *Bull. Astron. Inst. Czech.* 1, 13.
Link, F. (1950a). *Bull. Astron. (Paris)* **15**, 143.
Link, F. (1950b). *Trans. I.A.U.* **7**, 135.
Link, F. (1950c). *Bull. Astron. Inst. Czech.* **2**, 1.
Link, F. (1951). *Bull. Astron. Inst. Czech.* **2**, 131.
Link, F. (1952). *Bull. Astron. Inst. Czech.* **3**, 6.
Link, F. (1955a). *Meteorologie* **12**, 257.
Link, F. (1955b). "I.A.U. Symposium on Radioastronomy," No. 4, p. 400.

Link, F. (1956a). *Bull. Astron. Inst. Czech.* **7**, 1.
Link, F. (1956b). *Publ. Obs. Nat. (Prague).* No. 29.
Link, F. (1958). *Bull. Astron. Inst. Czech.* **9**, 169.
Link, F. (1959). *Bull. Astron. Inst. Czech.* **10**, 105.
Link, F. (1960a). *Bull. Astron. Inst. Czech.* **11**, 13.
Link, F. (1960b). *Proc. Nat. Conf. Acrosols, Liblice, Czechoslovakia, 1960*, pp. 539–544.
Link, F. (1961). *Studies Geophys. Geodet.* **5**, 64.
Link, F. (1962). *Bull. Astron. Inst. Czech.* **13**, 1.
Link, F. (1963). *Advan. Astron. Astrophys.* **2**, 101.
Link, F. (1966). *Bull. Astron. Inst. Czech.* **17**, 161.
Link, F., and Guth, V. (1936). *J. Observateurs* **19**, 129.
Link, F., and Guth, V. (1940). *Z. Astrophys.* **20**, 1.
Link, F., and Linková, Z. (1954a). *Publ. Obs. Nat. (Prague)* No. 25.
Link, F., and Linková, Z. (1954b). *Bull. Astron. Inst. Czech.* **5**, 82.
Link, F., Neužil, L., and Zacharov, I. (1967). *Ann. Geophys.* **23**, 207.
Linke, F. (1943). *Handbuch Geophys.* [8] **6**, 256.
Machin, K. E., and Smith, F. G. (1951). *Nature* **168**, 589.
Maedler, J. H. (1800). *Astron. Nachr.* **15**, 1.
Matsushima, S. (1966a). *Astron. J.* **71**, 699.
Matsushima, S. (1966b). *Univ. Iowa* **54**.
Matsushima, S. (1966c). *Nature* **211**, 1027.
Matsushima, S., and Zink, J. R. (1964). *Astron. J.* **69**, 481.
Matsushima, S., Zink, J. R., and Hansen, J. E. (1966). *Astron. J.* **71**, 103.
Maunder, E. W. (1921). *J. Brit. Astron. Assoc.* **31**, 346.
Mitra, S. K. (1952). "The Upper Atmosphere," p. 152. Calcutta.
Moore, R. C., and Schilling, G. F. (1965). Rand Corp. Rept. P-3130.
Moreno, H., and Stock, J. (1964). *Publ. Astron. Soc. Pacif.* **76**, 55.
Müller, G. (1883). *Publ. Obs. Potsdam* **3**, 285.
Oppolzer, T. (1887). *Denkschr. Akad. Wiss. Wien* **52**.
Paetzold, H. K. (1950). *Z. Naturforsch.* **5a**, 661.
Paetzold, H. K. (1951). *Z. Naturforsch.* **6a**, 639.
Paetzold, H. K. (1952). *Z. Naturforsch.* **7a**, 325.
Paetzold, H. K. (1953). *Z. Astrophys.* **32**, 303.
Palmén, E. (1934). *Meteorol. Z.* **51**, 134.
Pannekoek, A. (1903). *Astron. Nachr.* **164**, No. 3913.
Pettit, E. (1933). *Bull. Sol. Phenom. Zurich.* No. 24.
Pettit, E. (1938). *Bull. Sol. Phenom. Zurich.* No. 44.
Pingré, P. (1901). "Annales célestes." Bigourdan, Paris.
Prantl, E. (1881). "Aristotelis de Coelo," Libri IV, p. 21. Lipsiae.
Proctor, R. A., and Raynard, A. G. (1892). "Old and New Astronomy," pp. 503–509. London.
Przybylski, A. (1964). *Acta Astron. (Cracoriae)* **14**, 285.
Rakos, K. D. (1966). *Lowell Obs. Bull.* No. 131.
Ramanathan, R. R. (1929). *Nature* **123**, 834.
Reuss, J. D. (1804). "Repertorium Comentationum," 5. Goetingae.
Ricciolo, J. B. (1651). "Almagestum Novum." Bononiae.
Richardson, R. A., and Hulburt, E. O. (1949). *J. Geophys. Res.* **54**, 215.
Rocket Panel (1952). *Phys. Rev.* **88**, 1027.
Rougier, G. (1933). *Ann. Obs. Strasbourg* **2**, 3.
Rougier, G., and Dubois, J. (1944). *Ciel et Terre* **55**, No. 5.

Sanduleak, N., and Stock, J. (1965). *Publ. Astron. Soc. Pacific* **77**, 237.
Saussure, M. (1931). *Verhandl. Naturforsch. Ges. Basel* **42**, 391.
Schilling, G. F., and Moore, R. C. (1966). *Lunar Intern. Lab. Symp., Ist, Athens, 1965.* Springer, Berlin.
Schmidt, J. (1856). "Der Mond," p. 141. Leipzig.
Seeliger, H. (1896). *Abh. Bayer. Akad. Wiss., II Kl.* **19**, Part 2.
Sidran, M. (1968). *Advan. Astron. Astrophys.* **6**, 279–307.
Smith, M. (1885). *Phil. Mag. (I)* **66**, 169.
Snyder, C. W., Neugebauer, M., and Rao, U. R. (1963). *J. Geophys. Res.* **68**, 6363.
Struyck, N. (1740). "Inleiding to Algemeene Geographie." Amsterdam.
Surveyor III (1967). *NASA* **SP-146.**
Švestka, Z. (1948). *Bull. Astron. Inst. Czech.* **1**, 48.
Švestka, Z. (1949). *Bull. Astron. Inst. Czech.* **1**, 109.
Švestka, Z. (1950). *Bull. Astron. Inst. Czech.* **2**, 41.
Symons, G. L. (1888). Rept. Krakatoa Comm., London.
Tsesevich, V. P. (1940). *Publ. Astron. Inst. Univ. Leningrad* No. 50.
Van Diggelen, J. (1965). *Planetary Space Sci.* **13**, 271.
Vassy, E. (1958). *J. Sci. Meteorol.* **8**, 1.
Vigroux, E. (1954). *Ann. Astrophys.* **17**, 399.
Walker, M. F., and Reaves G. (1957). *Publ. Astron. Soc. Pacific.* **69**, 153.
Wirtz, C. (1915). *Astron. Nachr.* **201**, 22.
Witkowski, J. (1949). *Bull. Soc. Amis. Sci. Poznan (B)* **9**, 147.
Wolf, R. (1877a). "Geschichte der Astronomie," p. 174. München.
Wolf, R. (1877b). "Geschichte der Astronomie," p. 154. München.
Zacharov, I. (1952). *Bull. Astron. Inst. Czech.* **3**, 82.
Zacharov, I. (1961). Thesis, Prague.
Zwicky, F. (1937). *Phys. Rev.* **51**, 290 and 679.

Infrared Observation on the Eclipsed Moon

R. W. SHORTHILL and J. M. SAARI†
Boeing Scientific Research Laboratories
Seattle, Washington

I

INTRODUCTION

In the past decade, infrared measurements of the lunar surface during eclipse and lunation cooling have shown that the Moon exhibits a considerable degree of thermal heterogeneity. It was discovered that "thermal features" exist on the surface which, by reason of their slower cooling compared to their environs, are positive thermal anomalies, or "hot spots." The more prominent of these anomalies are associated with relatively young features, such as ray craters like Tycho or Aristarchus. Others are associated with rilles, older craters, regions of high albedo, or with regions where no particular feature can be identified.

In another chapter of this book, D. F. Winter reviews the infrared measurements and the status of the interpretation of these thermal features. From that discussion, it is clear that unambiguous explanations are not at hand for the thermal anomalies. In order to ascertain which of the several possible interpretations are correct, other sources of information besides the eclipse data will be helpful, including radar scattering data, lunation cooling on individual anomalies, lunation brightness data, high resolution photographs such as those obtained from the Lunar Orbiter spacecraft, and theoretical cooling curves for different models. Such studies, if successful in defining the physical properties responsible for the thermal anomalies, will then make possible the

† Deceased 12 January 1971.

determination of the geological significance of these features as related to the processes which have formed the lunar surface.

At the time of the present writing, the most extensive mapping of the lunar surface in the far infrared was accomplished during the 19 December 1964 eclipse (Saari and Shorthill, 1965; Shorthill and Saari, 1965). In the present chapter we shall confine ourselves to a brief description of these measurements and present several examples of the various methods used for analyzing the data. Further, the thermal contours obtained from one of the scan programs obtained during totality will be displayed. These contours have been overlaid onto the AIC charts which have been prepared for the so-called "Apollo Band," that is, the equatorial region $\pm 8°$ in latitude and $\pm 50°$ in longitude. For each of the charts a brief description of the more prominent thermal features is given including, in some instances, comparison with recent radar results obtained at 3.8 cm. Finally, some of the general characteristics of the anomalies in this band will be presented.

II
MEASUREMENTS

The instrumentation, described in detail previously (Shorthill and Saari, 1965; Saari and Shorthill, 1967), consisted of a mechanical scanning engine mounted at the Newtonian focus of the telescope. It was designed to scan the stationary lunar image in a rectangular raster. The separation between the individual scan lines was set to the aperture diameter (10″ arc) so that ~200 scan lines covered the lunar disk. The infrared emission was measured with a mercury-doped germanium photodetector whose response was restricted to 10–12 μ by an interference filter so as to minimize sky emission. The data were recorded on tape and subsequently digitized for computer reduction.

The eclipse measurements were made with the 74-in. Kottamia telescope of

TABLE I

Circumstances of the Total Lunar Eclipse of 19 December 1964

	h	m	
Moon enters penumbra	00	01.3	E. T.
Moon enters umbra	00	59.8	E. T.
Totality begins	02	07.8	E. T.
Middle of eclipse	02	37.9	E. T.
Total eclipse ends	03	07.9	E. T.
Moon leaves umbra	04	16.0	E. T.
Moon leaves penumbra	05	14.4	E. T.

the Helwan Observatory in Egypt (Samaha, 1967). In order to keep the lunar image fixed at the Newtonian focus, a guidescope mounted on the telescope was used for guiding on a small sharp lunar feature. Occasionally errors in guiding occurred, especially during totality when the Moon was dim, which caused the image to shift somewhat (several seconds of arc) at the focus. The circumstances of the eclipse are given in Table I. A single scan program required 17 min for completion; in all, nine scans were made through the eclipse including one on the full moon (Table II).

TABLE II

ECLIPSE SCANS FOR 19 DECEMBER 1964

Scan number	Mid-scan time (U.T.)	Phase angle Disk center	Zenith angle	Remarks
8 H	$23^h\ 34^m.28$	$-2°$ 16.12	25°.86	Full moon (Dec. 18)
9 H	0 19.00	-1 58.21	35.10	Partly in penumbra
10 H	0 41.05	-1 48.90	39.67	Partly in penumbra
11 H	1 06.42	— —	45.03	Penumbra
12 H	1 35.63	— —	50.93	Partly in umbra
13 H	2 18.82	— —	59.75	Umbra
14 H	2 42.05	— —	64.44	Umbra
15 H	3 02.82	— —	68.60	Umbra
16 H	3 48.70	— —	77.61	Partly in umbra

The calibration of the infrared detector was accomplished with respect to a computed temperature of the subsolar point on the pre-eclipse scan. The calculation included corrections for the Sun–Moon distance, the local albedo, and directional effects. On the eclipse scans corrections were made to the extinction coefficient taking into account the precipitable water in the atmosphere inferred from ground level humidity measurements.

III

EXPERIMENTAL RESULTS

Digitization of each scan program produced 140,000 data points so that over 10^6 data points were obtained from the nine relevant scans. A qualitative overview of data in this amount can be achieved by construction of images. A complete collection of images of the infrared and photometric data have been published (Saari *et al.*, 1966). In Fig. 1 we show, by way of example, scan 12 H during the penumbral phase. The west limb of the Moon is in the umbra; the bright feature near the bottom is the crater Tycho and Mare Humorum is the thermally
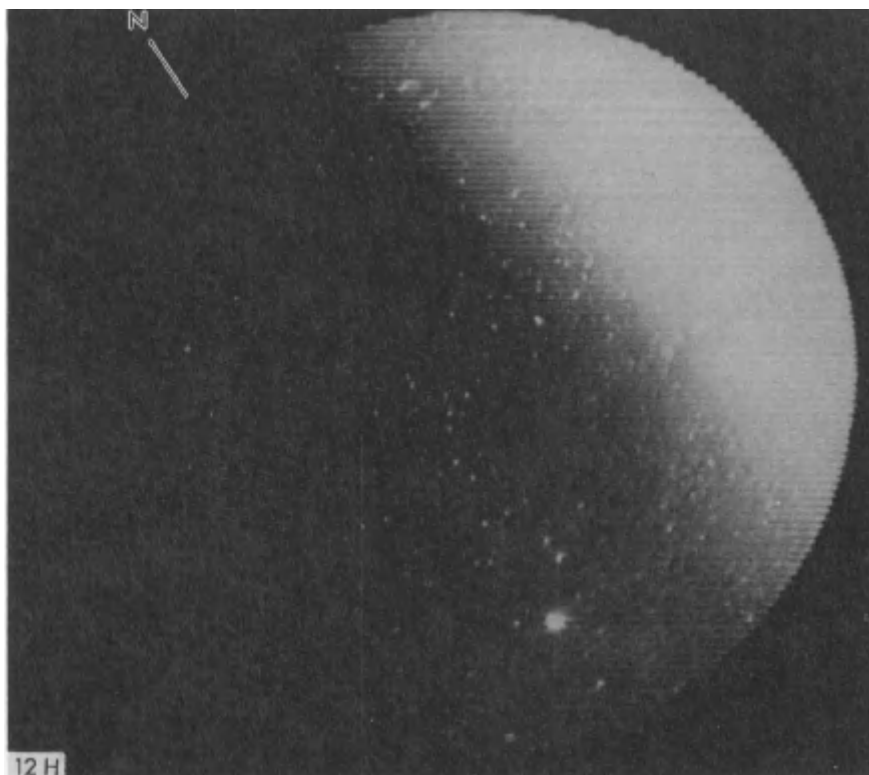
FIG. 1. Reconstructed 11 μ infrared image of the Moon. The east limb is still partly in the penumbra.

enhanced circular feature to the left of Tycho. Figure 2 shows scan 13 H, the first of three in totality, in which hundreds of hot spots and enhancements in some of the seas are evident. In the present discussion we shall be concerned primarily with observations within the Apollo Band.

Eclipse cooling curves are useful in inferring the thermophysical properties of the lunar surface. Examples are shown in Fig. 3 for two mare regions at the center of the Moon and the Surveyor V landing site in Mare Tranquillitatis. Another example is the cooling curve for a thermal anomaly in the ray system west of Kepler at 42.9°W, 7.1°N shown in Fig. 4. During totality, this feature was detected on two successive scan lines; in the figure the two peak temperatures are plotted, as well as the environ temperature. It is apparent from the figure that the peak temperature attributed to a thermal anomaly depends critically on its position in the raster pattern on successive scan programs. However, it is possible to make an "areal correction" if the size of the feature is
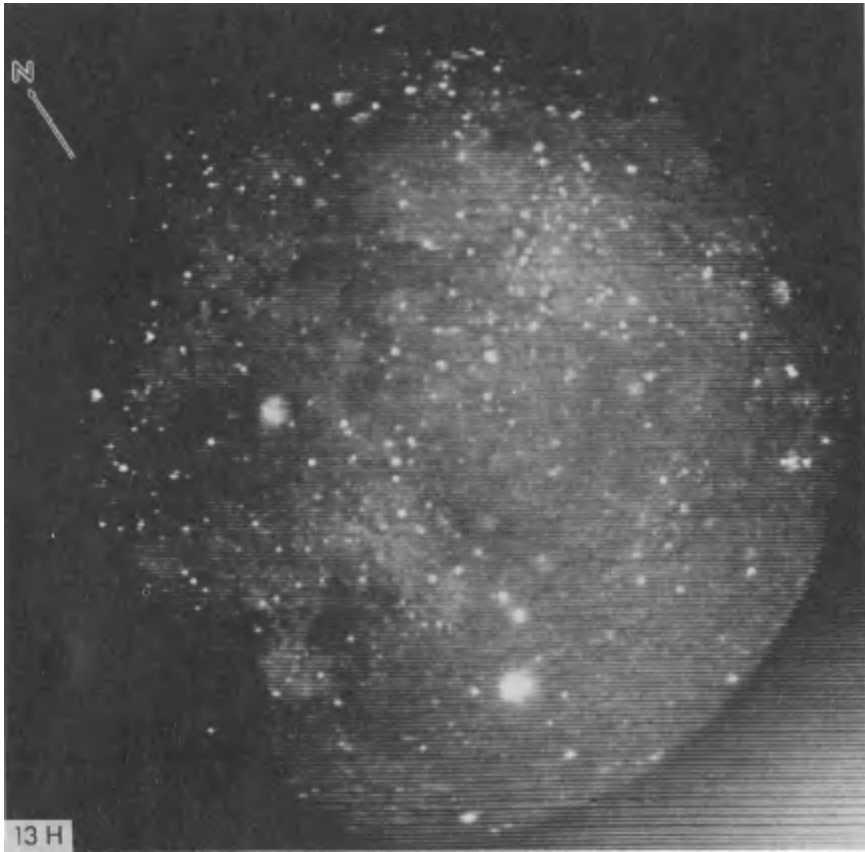
FIG. 2. Reconstructed 11 μ infrared image of the Moon in totality. Hundreds of hot spots and thermal enhancements in some of the mare are evident.

known. In the case of craters, the simplest assumption is that the size of the thermal anomaly is equal to that of the crater itself. On this basis, areal corrections have been performed on a number of craters (Shorthill and Saari, 1965). For the particular anomaly under consideration here, there is no sizeable crater at its position; accordingly, it was assigned (rather arbitrarily) a diameter of 18 km. Since the relative magnitude of the two signals on the successive scan lines is related to the geometry of intersection, it is possible to solve for the temperature of the feature. In the figure it is seen that the corrected temperature provides a smooth curve which parallels the environ data.

Detailed studies of the spatial distribution of anomalies require the construction of isothermal contours. By way of illustration, contours drawn from scan 15 H, the last scan program in totality, are shown in Figs. 5–10. The contour
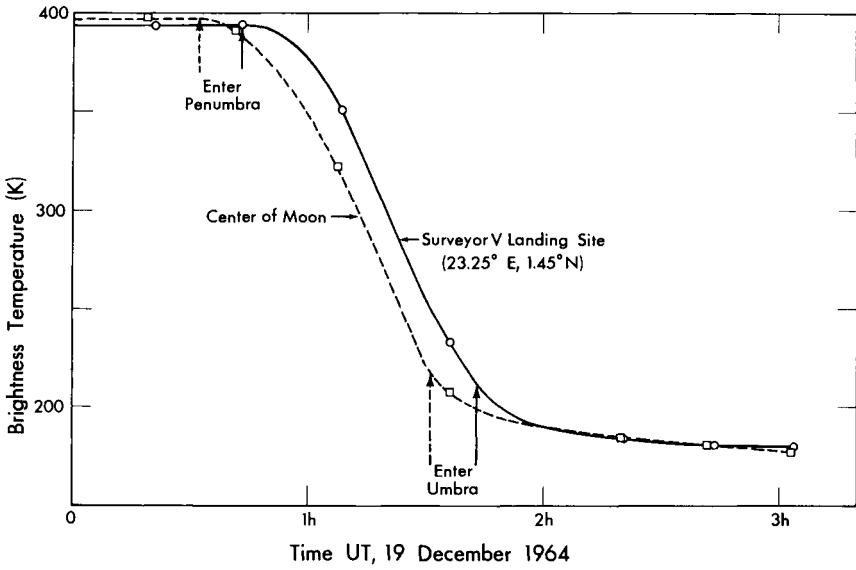
FIG. 3. Eclipse cooling curves for the center of the Moon and Surveyor V landing site in Mare Tranquillitatis.
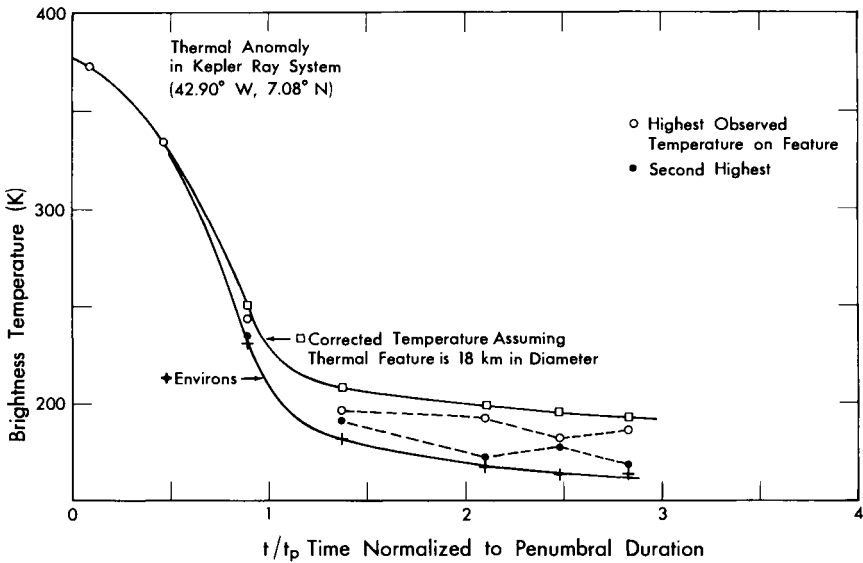
FIG. 4. Eclipse cooling curve for a hot spot (26172 X) in the ray system of Kepler. This is also the location of a radar enhancement (see description of Maestlin AIC 57 D).
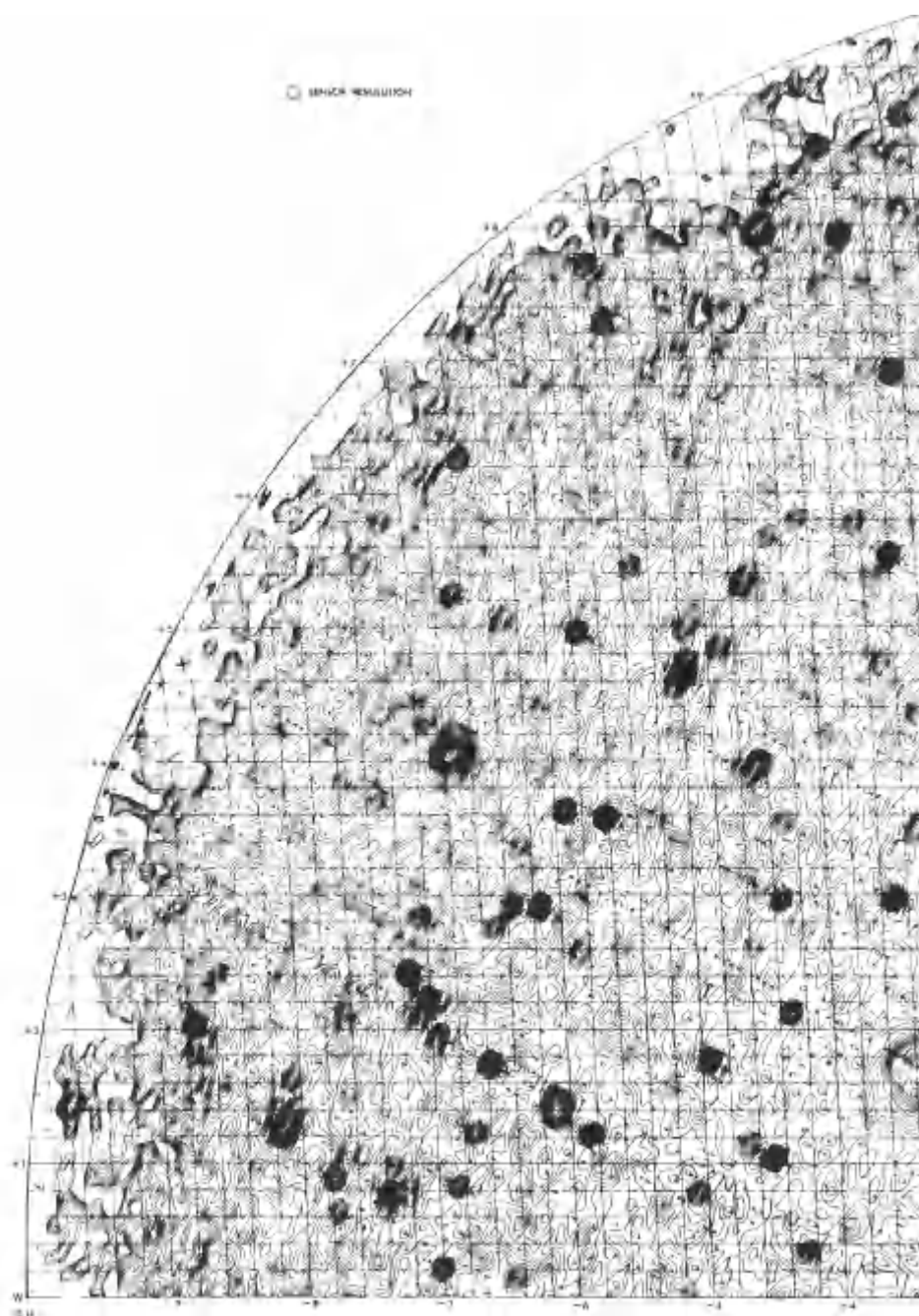
FIG. 5. Eclipse isothermal contours for scan 15 H. This covers the major portion of the II quadrant.
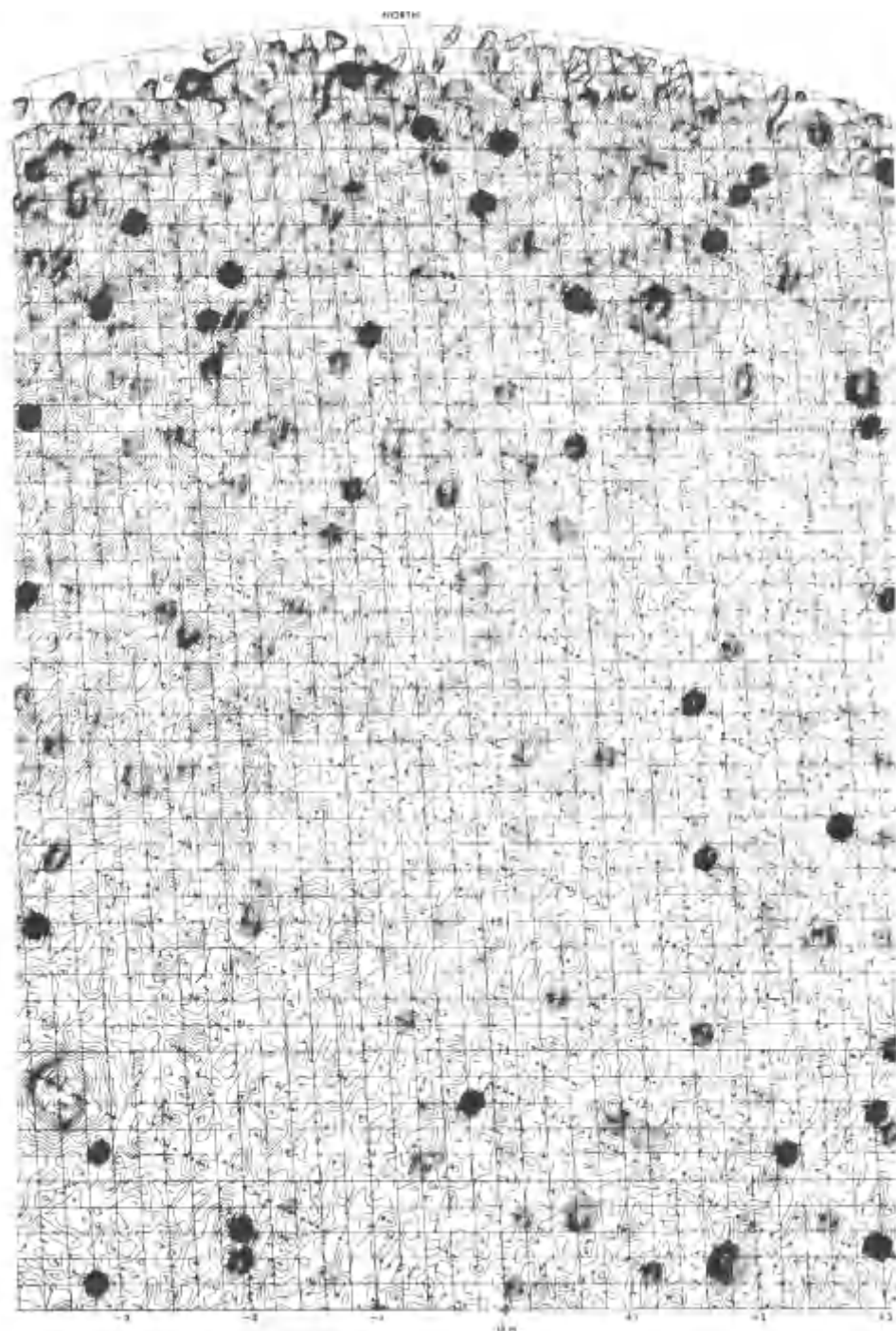
FIG. 6. Eclipse isothermal contours for scan 15 H. This covers part of the I and II quadrant.
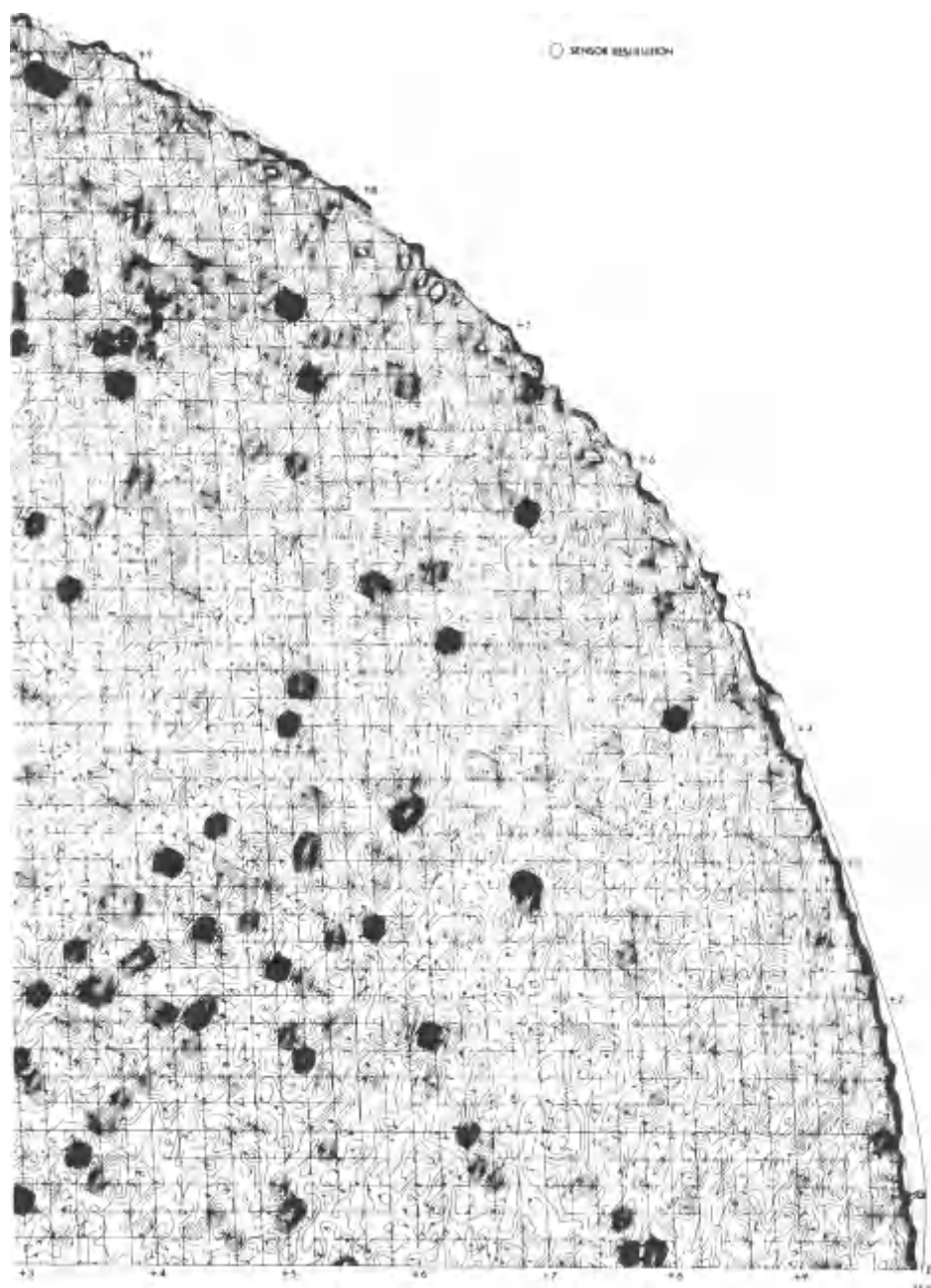
FIG. 7. Eclipse isothermal contours for scan 15 H. This covers the major portion of the I quadrant.

FIG. 8. Eclipse isothermal contours for scan 15 H. This covers the major portion of the III quadrant.
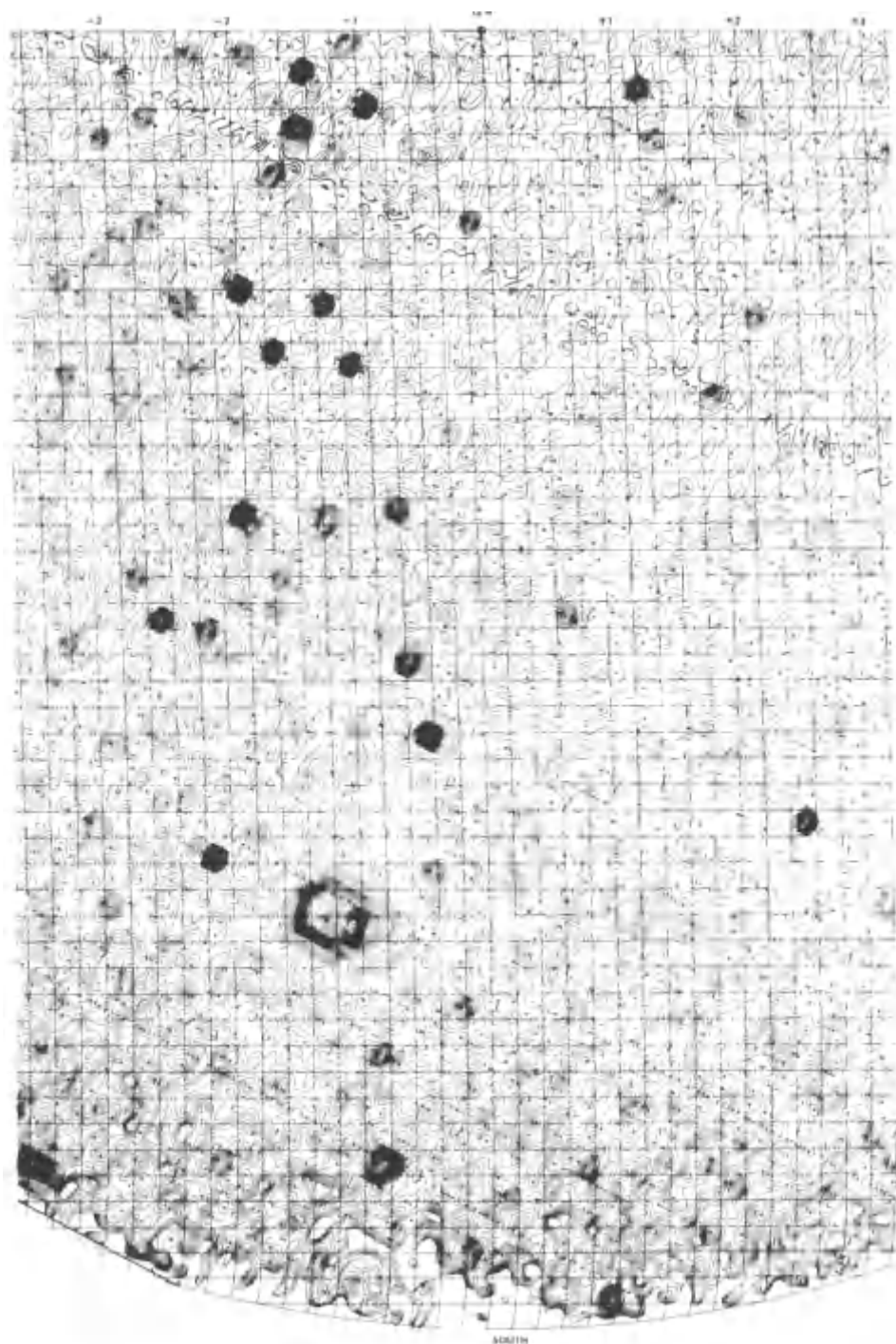
FIG. 9. Eclipse isothermal contours for scan 15 H. This covers part of the III and IV quadrant.
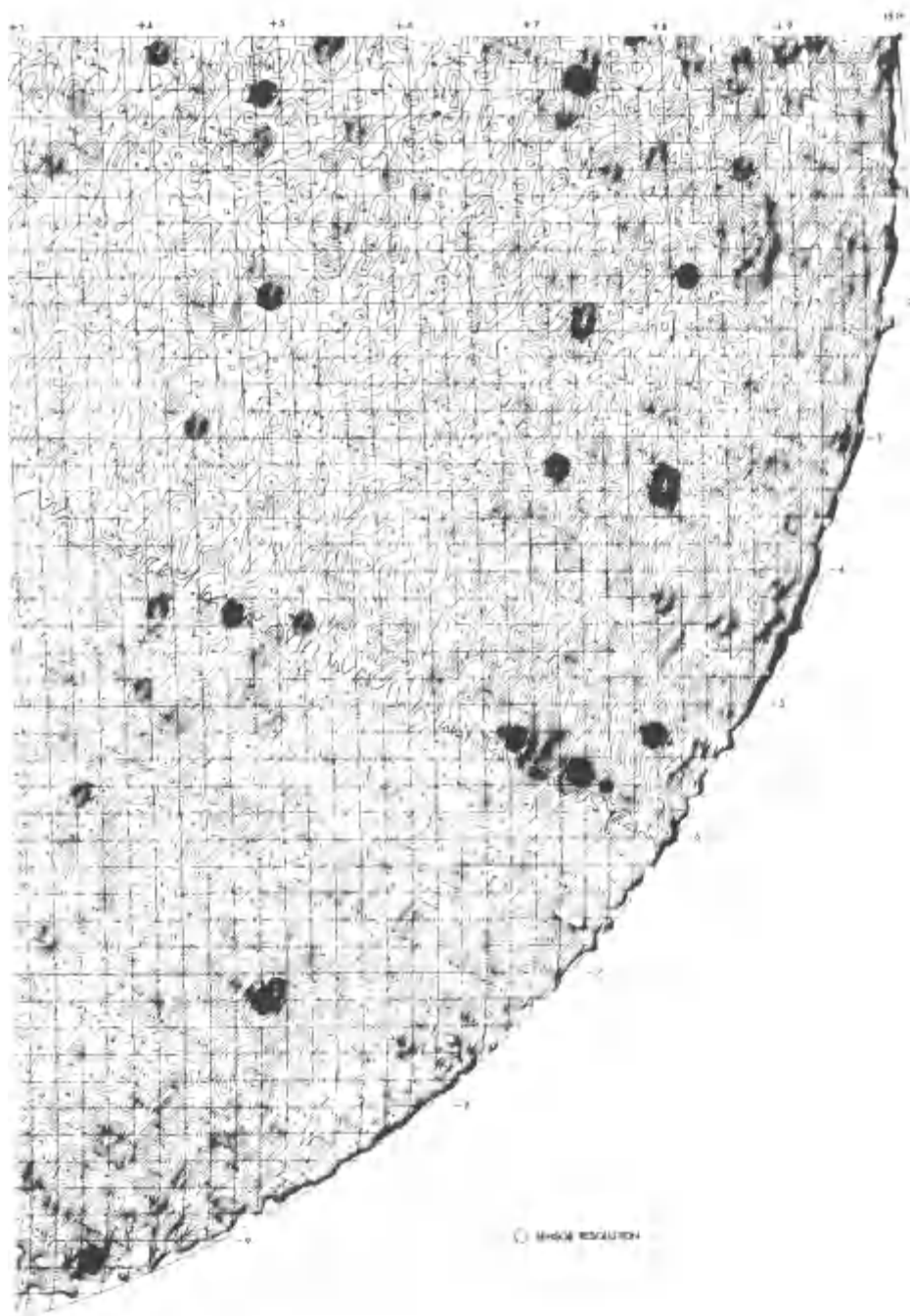
FIG. 10. Eclipse isothermal contours for scan 15 H. This covers the major portion of the IV quadrant.

interval is 2 K; the lowest temperature is 140 K corresponding to contour level #1 on the chart.

IV

METHOD FOR PRODUCING THERMAL OVERLAYS FOR THE AIC CHARTS

Twenty AIC charts of the Apollo Band have been drawn by the United States Air Force Aeronautical Chart and Information Center of St. Louis, Missouri. Each chart covers 10° in longitude and 8° in latitude from the equator and was originally drawn to a scale of 1:500,000. The isothermal contours for scan 15 H were manually transferred to the coordinate base for each AIC chart using the selenographic grid shown in Figs. 5 through 10. This grid was automatically drawn using a position fit obtained on 45 small thermal anomalies identifiable with specific craters whose selenographic position is known. Specifically, the positions of these features were measured in the scanner coordinates. Then a least square fit was performed with reference to their positions on a selenographic grid. The resulting transformation from selenographic to scanner coordinates made it possible to draw automatically the longitude–latitude grid shown superimposed over the contours.

The formal rms error resulting from the above transformation on the 45 features amounted to 2.3″ arc or approximately one-quarter of the resolution element. This is a measure of the position error of the contours. However, as will become apparent in the discussion of the individual charts, in certain regions systematic errors are present which arose, for instance, from guiding errors. In addition, errors of ~0.2° in selenographic position can arise in the process of manually transferring the contours to the AIC base.

V

REMARKS

The thermal overlays for the twenty AIC charts are shown in Figs. 11–30. Maximum and minimum temperatures indicated on thermal features were obtained from the original digital data. Hachure marks show relative lows. The temperature levels have been labeled on a number of the contours.

Table III lists 165 thermal anomalies found in the region covered by these charts (only positive anomalies have been cataloged, and for the most part only those with a temperature difference $\Delta T > 5\,K$ over their environs). In the catalog, the Lunar and Planetary Laboratory number is the designation of the feature in terms of its standard coordinates (ξ, η) listed in the Lunar and
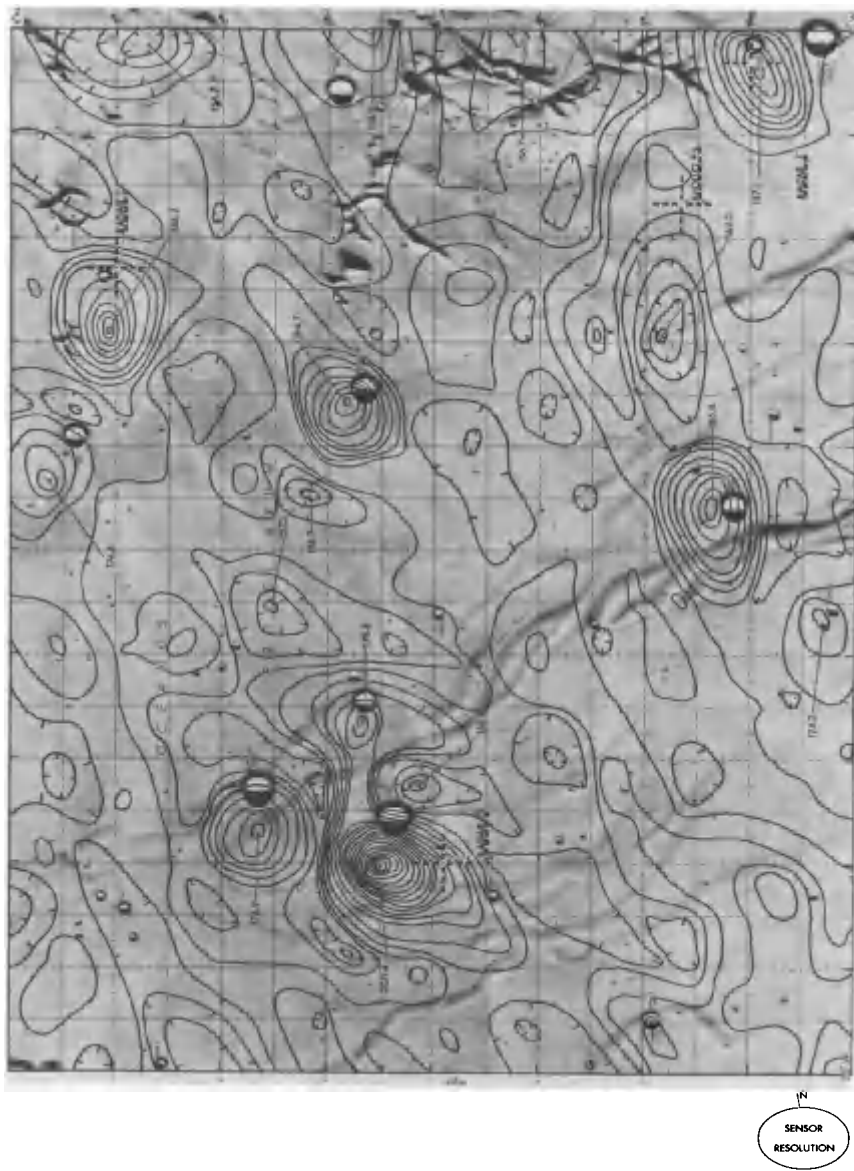
FIG. 11. Maestlin AIC 57 D 50°–40°W; 0°–8°N. The projected size and orientation of the sensor is shown for the center of each AIC chart.

Fig. 12. Encke AIC 57 C 40°–30°W; 0°–8°N.

Fig. 13. Reinhold AIC 58 D 30°–20°W; 0°–8°N.

FIG. 14. Gambart AIC 58 C 20°–10°W; 0°–8°N.
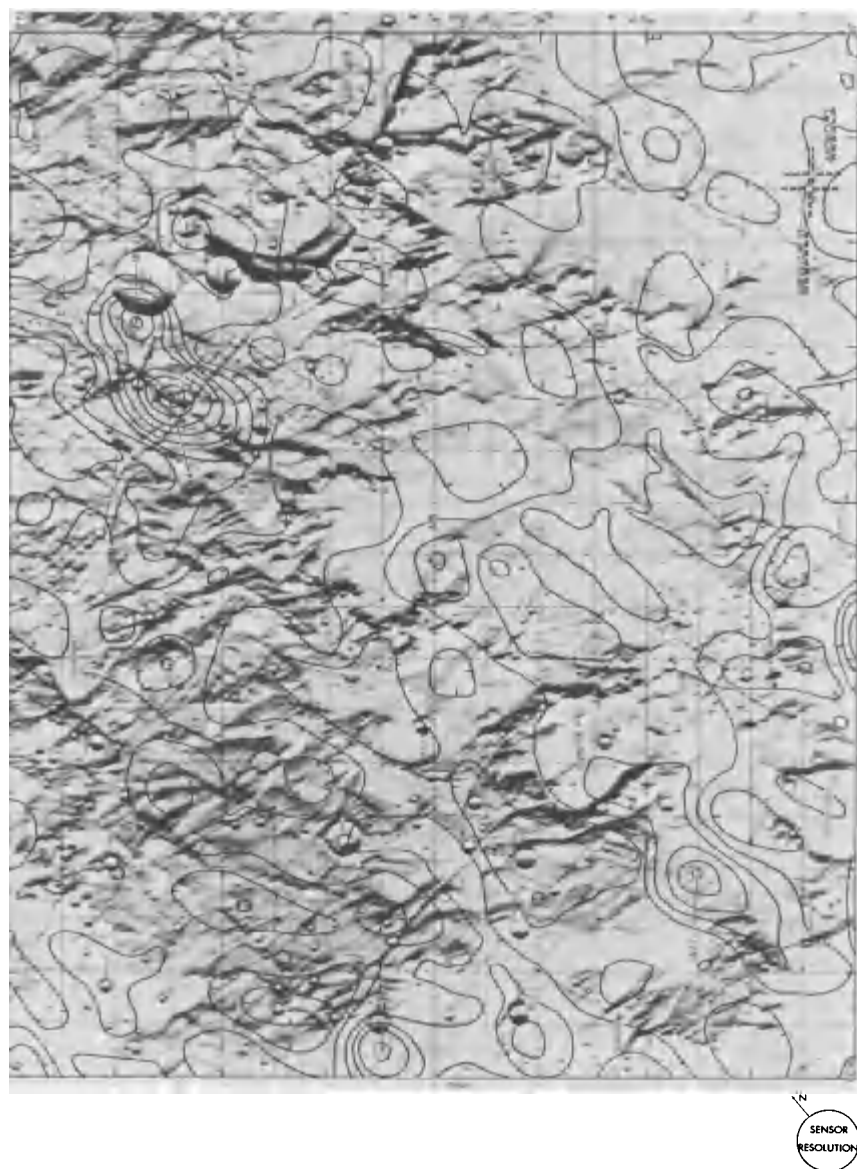
FIG. 15. Pallas AIC 59 D 10°–0°W; 0°–8°N.

FIG. 16. Triesnecker AIC 59 C 0°–10°E; 0°–8°N.
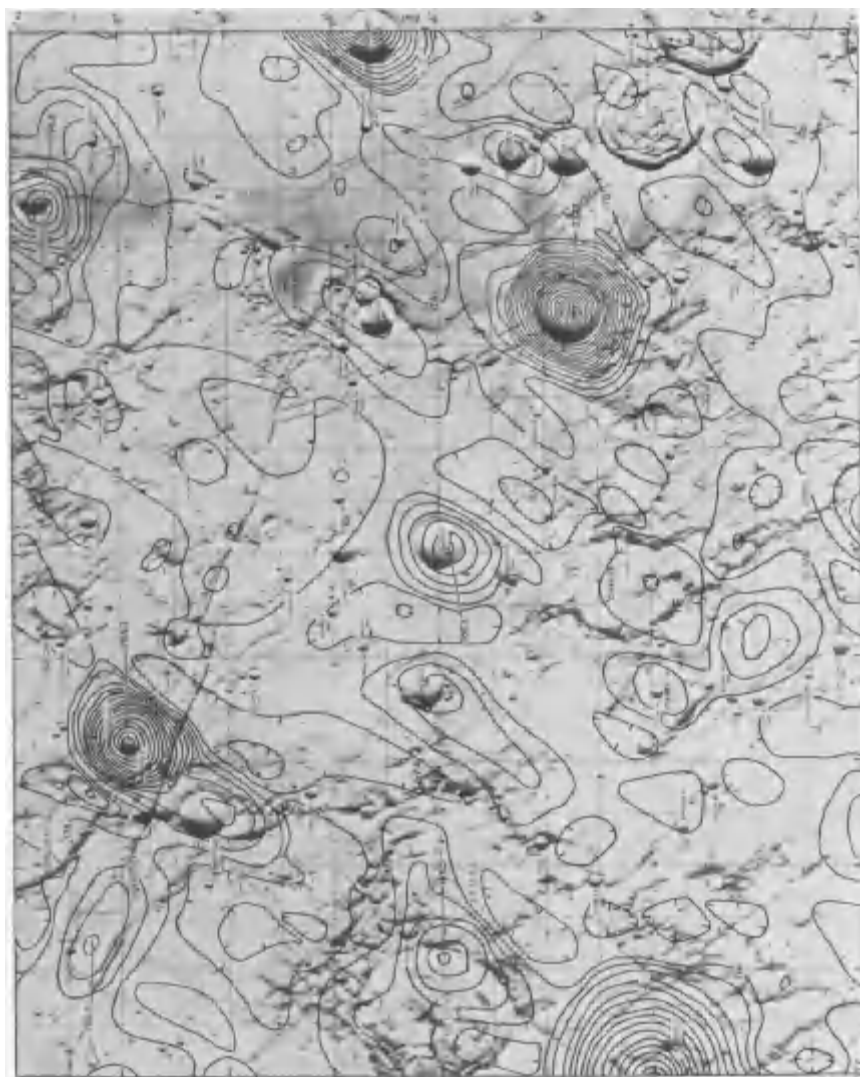
Fig. 17. Agrippa AIC 60 D 10°–20°E; 0°–8°N.
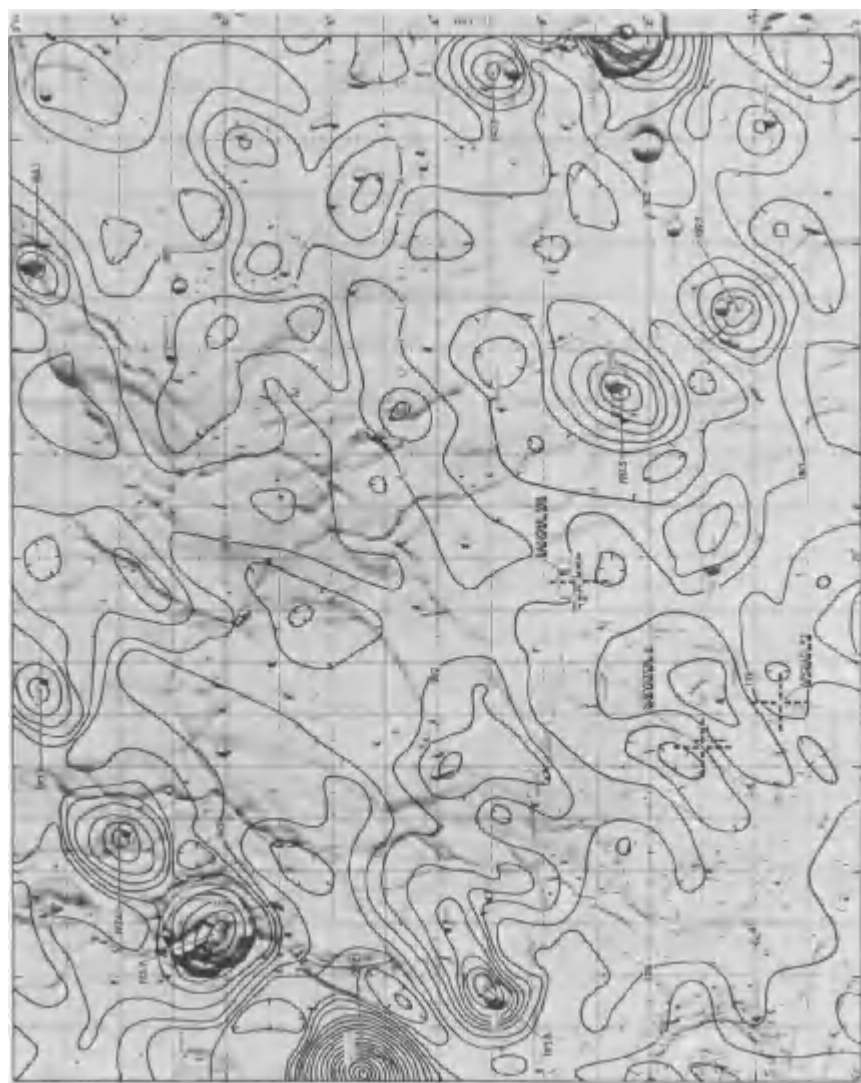
FIG. 18. Arago AIC 60 C 20°–30°E; 0°–8°N.
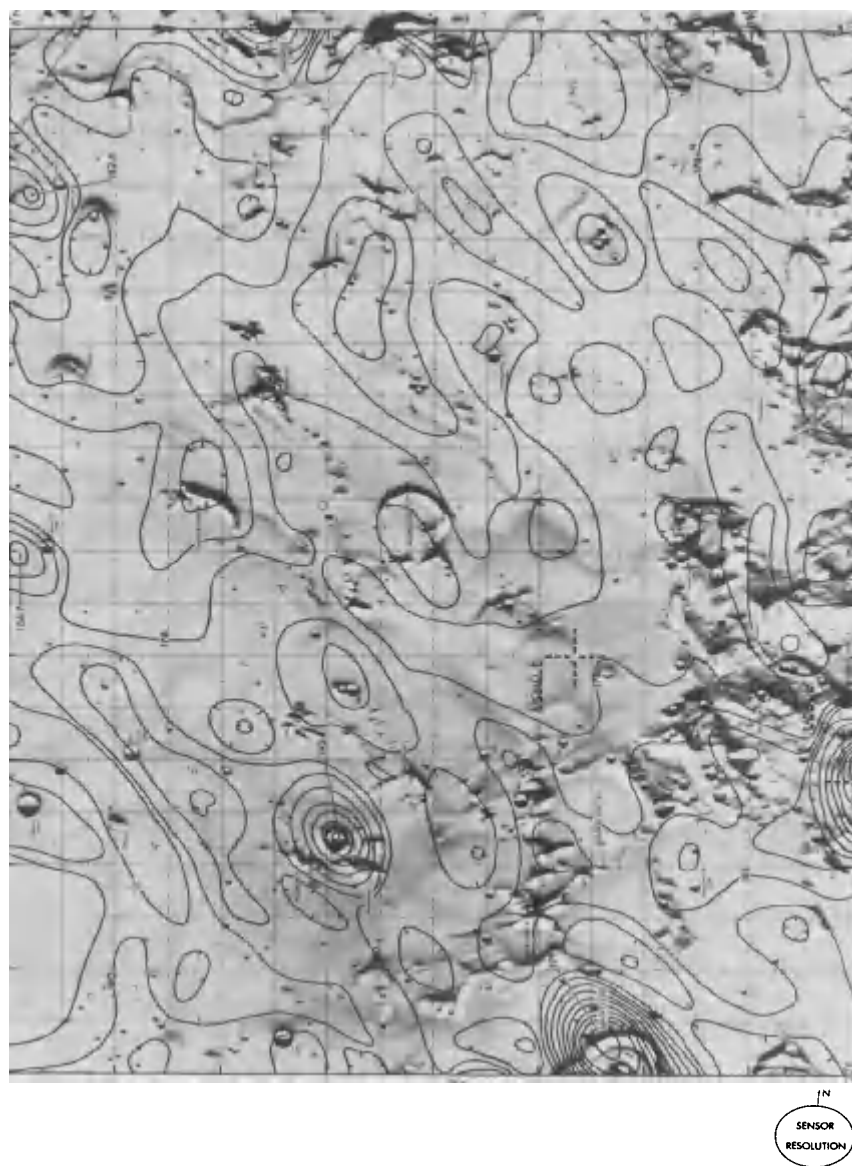
FIG. 19. Maskelyne AIC 61 D 30°–40°E; 0°–8°N.

FIG. 20. Secchi AIC 61 C 40°–50°E; 0°–8°N.

N
SENSOR RESOLUTION
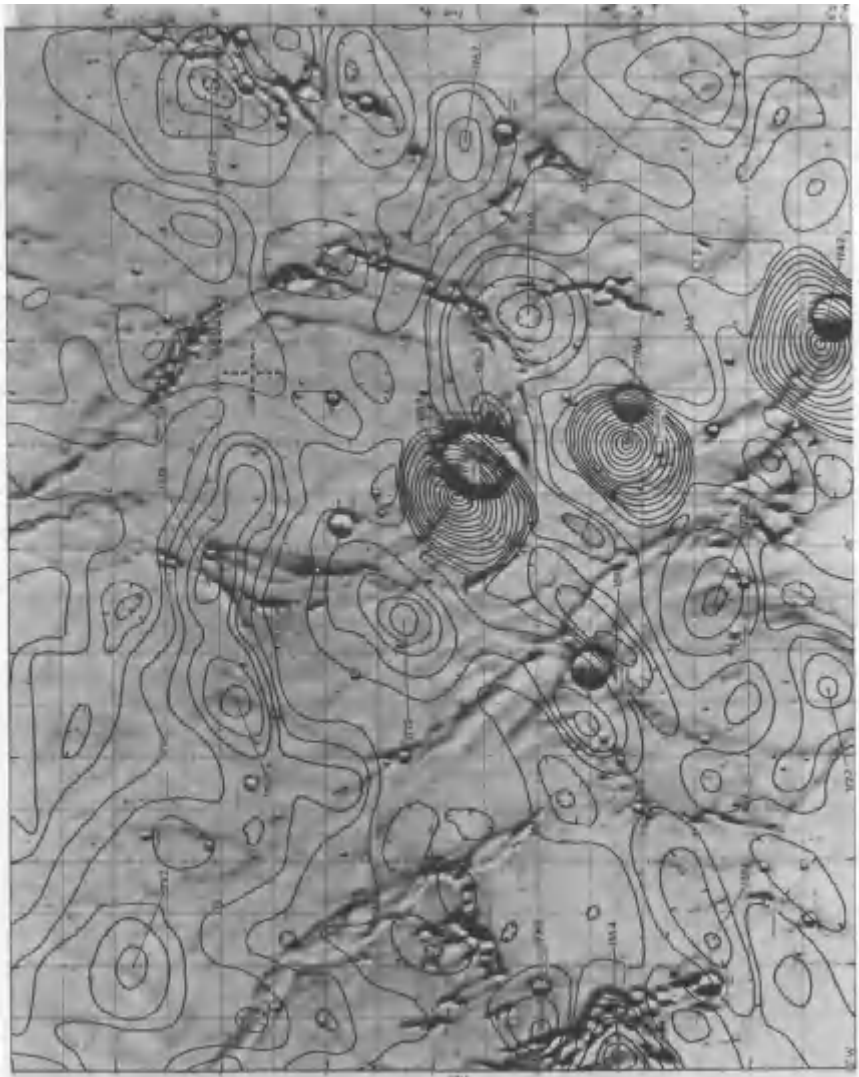
Fig. 21. Flamsteed AIC 75 A 50°–40°W; 0°–8°S.
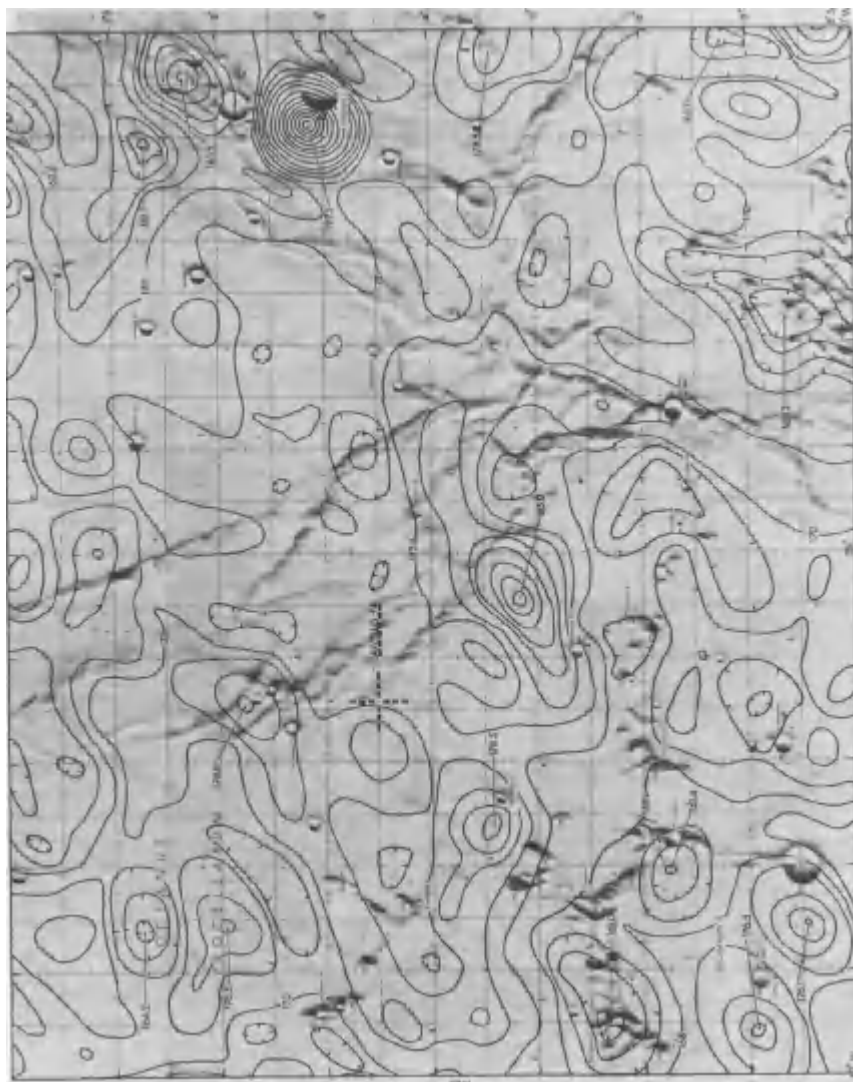
FIG. 22. Wichmann AIC 75 B 40°–30°W; 0°–8°S.
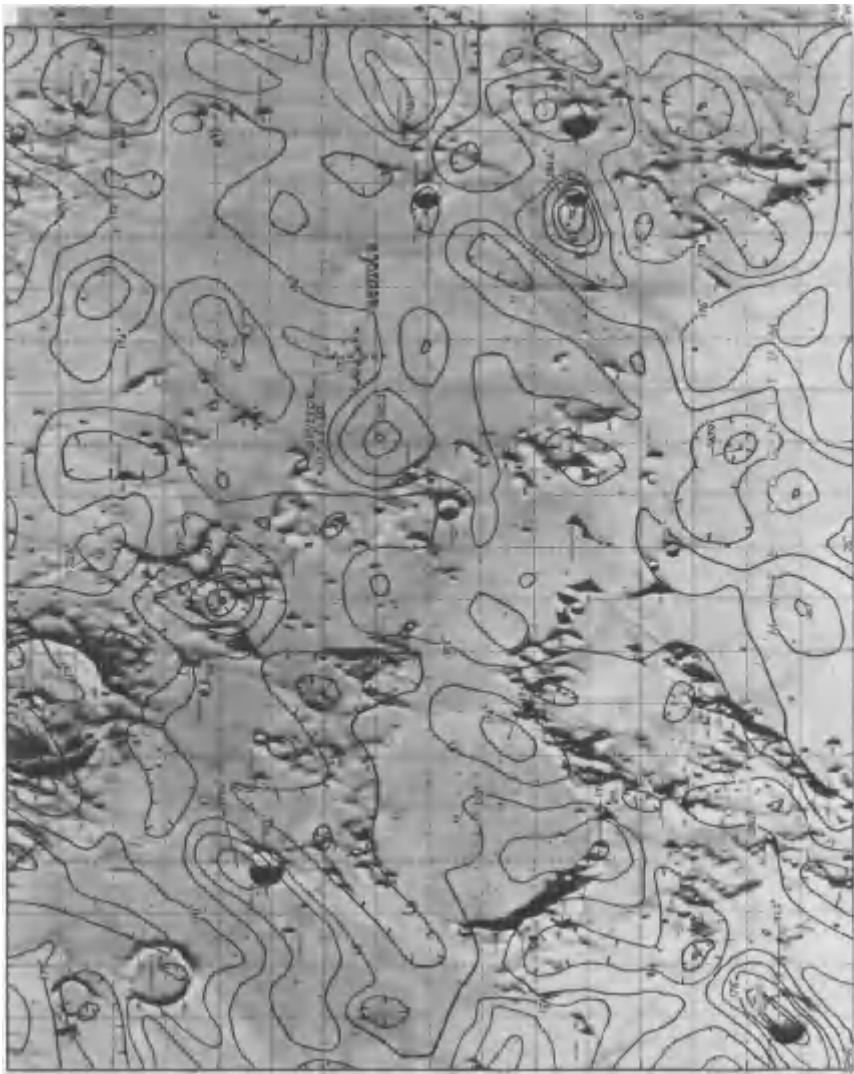
Fig. 23. Euclides P AIC 76 A 30°–20°W; 0°–8°S.

FIG. 24. Fra Mauro AIC 76 B 20°–10°W; 0°–8°S.
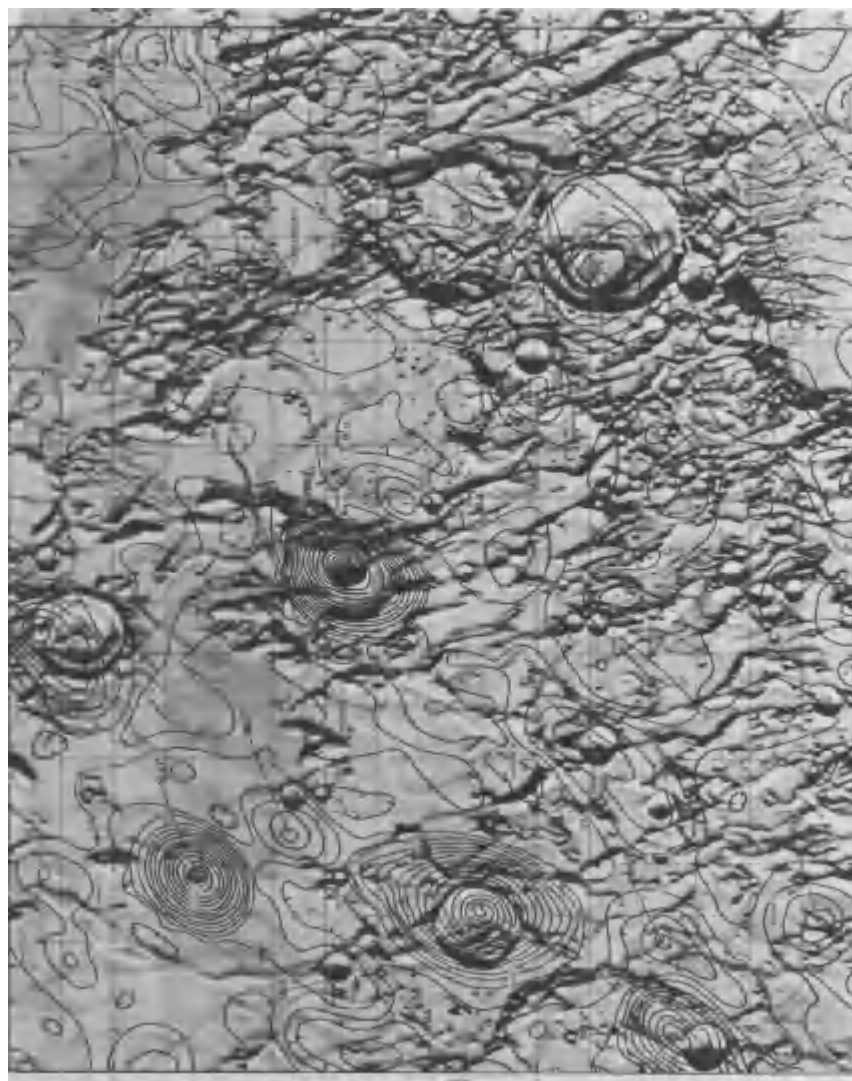
Fig. 25. Flammarion AIC 77 A 10°–0°W; 0°–8°S.
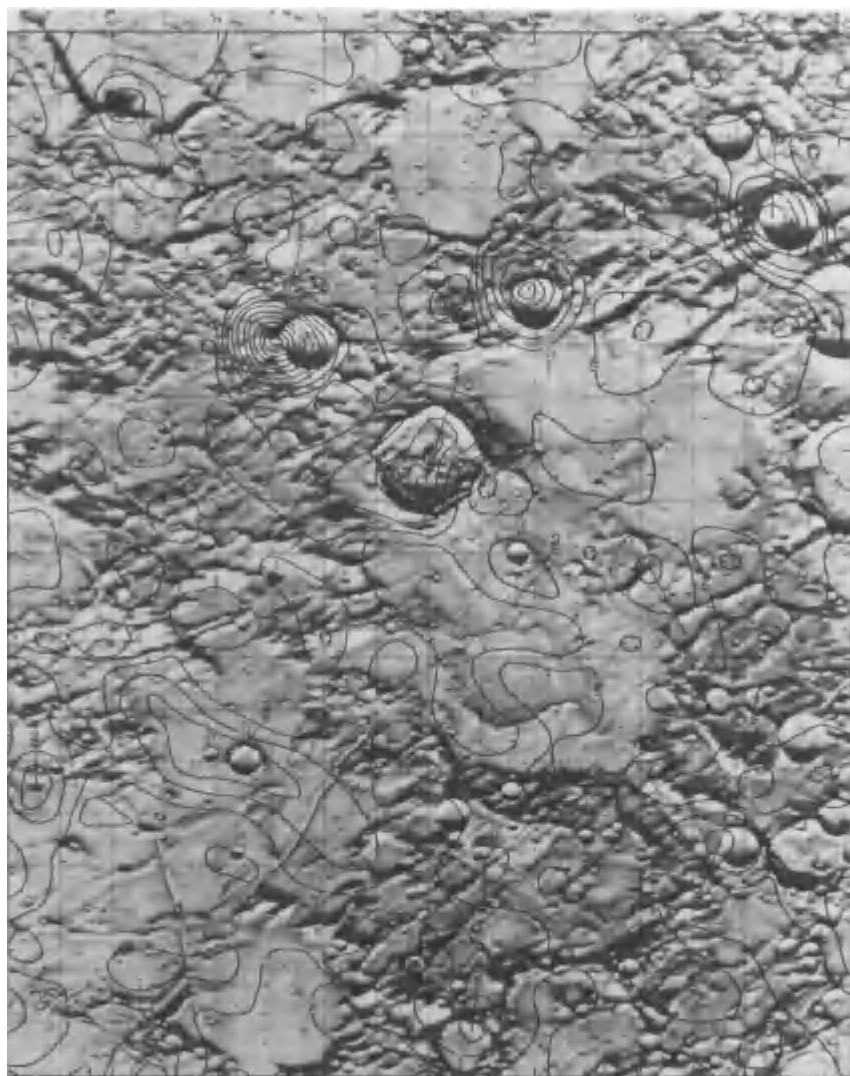
FIG. 26. Hipparchus AIC 77 B 0°–10°E; 0°–8°S.
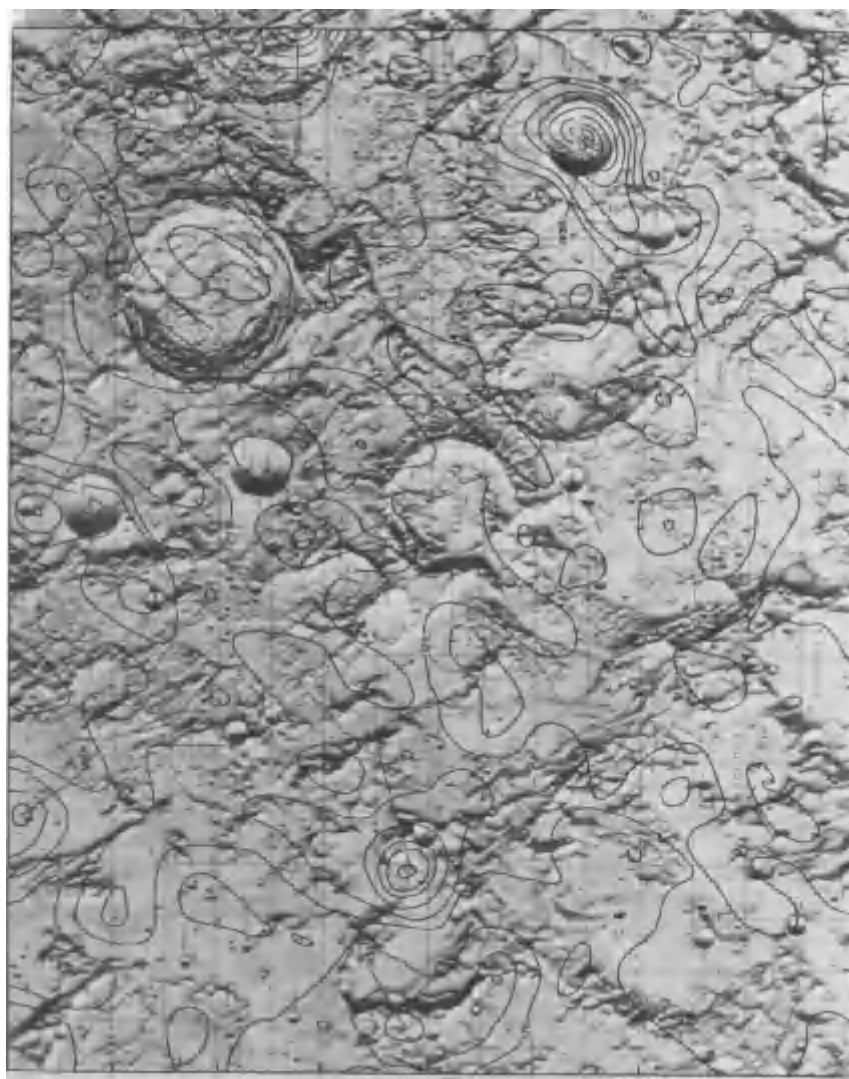
FIG. 27. Delambre AIC 78 A 10°–20°E; 0°–8°S.

FIG. 28. Torricelli AIC 78 B 20°–30°E; 0°–8°S.

N
SENSOR
RESOLUTION

Fig. 29. Capella AIC 79 A 30°–40°E; 0°–8°S.

FIG. 30. Messier AIC 79 B 40°–50°E; 0°–8°S.

TABLE III

CATALOG OF THERMAL ANOMALIES IN THE APOLLO BAND

L & PL Number	Designation	Coordinates λ long.	β lat.	Diam (km)	C	B	CE	ΔT (K)	LO IV	Remarks
10002	Bruce	0.4 E	1.1 N	6.69	1	pM	0	13.3	102	
10016	Chladni	1.1	4.0	13.61	1	pM	0	12.2	102	
10022	Blagg	1.4	1.2	5.41	1	pM	0	8.7	102	
10067	Triesnecker	3.6	4.1	25.95	1	pM	PP	17.9	102	
10102	Pallas	0.5	7.0	5.74	1	C	0	8.9	102	
11006	Triesnecker D	5.9	3.4	6.05	1	C	0	6.2	097, 108	
11012	Rhaeticus B	6.8	1.6	5.82	1	C	0	14.8	097	
11073	Godin	10.2	1.8	34.76	2	C	P	33.0	097	
11097 R	Rim (Agrippa)	11.2	4.0	(46.04	2	C	P)	10.5	097	
11103	Hyginus	6.3	7.7	10.65	3f	pM	0	11.9	097	Follows rille
11192 X	No Feature	11.3	7.3	—	—	C ?	—	7.0	097	Rima Ariadaeus(?)
12037	Whewell	13.7	4.1	13.61	2	C	0	5.0	090	
12041 X	No Feature	14.2	1.0	—	—	—	—	7.0	090	Moderately rough
12066	Cayley	15.1	3.9	14.32	1	C	0	11.3	090	Cool halo
12094	Dionysius	17.2	2.8	17.64	1	C	0	49.3	090	
12097	Ariadaeus	17.2	4.5	11.16	2	pMC	0	5.0	090	
12110	Silberschlag	12.5	6.2	13.42	1	C	0	7.0	090	
12122	Silberschlag A	13.2	6.9	6.88	1	C	0	28.3	090	
13025	Ritter B	18.9	3.2	13.92	2	aM	0	7.0	085	
13047	Manners	20.0	4.5	15.14	1f	pM	0	29.1	085	
13056	Arago B	20.8	3.4	6.55	1	pM	0	15.8	085	
13113	Sosigenes A	18.4	7.7	11.87	1	pM	0	16.4	090	
13172	Arago D	22.3	6.8	4.55	1	pM	0	15.4	085	
13193	Arago EA	23.6	7.8	~4.0	1	pM	0	11.1	085	
14044	Maskelyne G	26.7 E	2.2 N	6.45	1	pM	0	12.5	078	Cool halo

14047 X	No Name Crater	26.4 E	4.3 N	<2.0	1 ?	pM ?	0 ?	4+	078	Moderately rough
14062	Maskelyne X	27.3	1.3	4.21	1	pM	0	9.9	118	
14095	Maskelyne K	29.6	3.2	5.28	1	pM	0	8.8	078	
14163	Maskelyne M	27.8	7.8	7.63	1	pM	0	6.1	078	
15003	Maskelyne	30.0	2.1	23.76	2	pM	P	17.6	078	
15038	Maskelyne H	32.2	4.9	6.48	1	pM	0	12.4	073, 078	
15153 N	Near (Cauchy M)	34.0	7.92	(4.71	1	pM	0)	6.7	073	
16046	Taruntius F	40.5	3.9	10.74	1	pM	0	15.7	066	
16049	Taruntius E	40.2	5.5	11.28	1	pM	0	21.5	066	
16059	Taruntius EB	41.2	5.2	4.69	1	pM	0	7.1	066	
16066	Secchi B	41.6	3.8	5.25	1	pM	0	14.8	066	Radar "0"
16113 N	Near (Cauchy CA)	38.4	7.8	(<4.0	1	pM	0)	8.4	066	
17036	Taruntius TB	47.4	3.6	~6.5	1	pM	0	9.8	061	
17053	Taruntius G	49.4	1.8	9.92	1	pM	0	20.3	061	
17060	Taruntius H	49.8	0.3	8.45	1	pM	0	29.8	061	
17110	Taruntius C	45.9 E	6.2 N	10.81	1	pMC	0	8.4	061	
20131 X	No Name Crater	1.8 W	6.3 N	~7.0	2 ?	C ?	0 ?	5.1	109	Moderately rough
20141 R	Rim (Bode)	(2.4	6.7	18.58	1	p	R)	10.2	109	
20161	Bode G	3.5	6.3	4.40	1	C	0	14.4	109	
21042 X	No Name Crater	8.1	1.5	—	—	—	—	~9.0	114, 109	
21067	Schröter D	9.5	4.5	5.02	1	pM	0	8.6	114	
22003	Gambart B	11.5	2.2	11.49	1	pM	0	22.1	114	
22005	Gambart C	11.8	3.3	12.18	1	pM	0	35.0	114	
22101 X	No Feature	11.8	6.3	—	—	pM ?	—	~5.0	114	
23021	Gambart A	18.7	1.0	11.98	1	pM	0	29.7	121	
23040	No Feature	20.2	0.02	—	—	—	—	~5.0	121	
23085 R	Rim (Reinhold)	23.45	2.92	(47.47	1	pM	pp)	9.0	126	On rim
23112	Copernicus H	18.3	6.9	4.62	1	pM	0	20.3	121	Cool halo
23173 N	Near (Copernicus B)	22.5	7.75	—	—	—	—	~5.0	126	Moderately rough
24018 X	No Feature	24.3 W	4.8 N	—	—	—	—	5+	126	Radar enhancement

(*continued*)

TABLE III (*continued*)

L & PL		Coordinates		Diam				ΔT		
Number	Designation	λ long.	β lat.	(km)	C	B	CE	(K)	LO IV	Remarks
24022	Reinhold NA	24.5 W	1.9 N	2.24	1	pM	0	20.0	126	Radar "L"
24140	Hortensius C	26.7	5.9	6.74	1	pM	0	27.7	126	
24161 R	Rim (Hortensius)	28.6	6.5	(14.65	1	pM	0)	11.3	126, 133	Cool halo
25007	Hortensius A	30.7	4.4	10.15	1	pM	0	17.8	133	Cool halo
25010	Lansberg A	31.1	0.2	8.62	1	pM	0	7.4	132, 133	Cool halo
25041 X	No Feature	33.1	0.7	—	—	—	—	~6.5	132	
25091	Encke C	36.4	0.6	8.46	1	pM	0	6.6	138	
25094	Encke B	36.8	2.3	11.47	1	pM	0	5.5	138	
25182	Kepler A	36.1	7.1	10.99	1	pMC	0	23.2	138	
26041	Encke X	40.2	0.9	3.53	1	pM	0	17.1	138, 144	Radar "J"
26048 X	No Feature	40.0	4.67	—	—	—	—	8.3	144	
26074 X	No Feature	42.8	2.4	—	—	—	—	5.0	144	Cool halo
26088	Maestlin H	40.6	4.9	7.13	1	pM	0	18.7	144	
26104	Kepler	38.0	8.1	31.55	2	pM	pp	27.1	138	
26172 X	No Feature	42.9	7.1	—	—	—	—	18.2	144	Radar "I"
26193	No Name Crater	44.3	7.6	2.0	1	—	0	8.4	144	
27002	Suess F	44.6	1.1	7.63	1	pM	0	17.8	144	
27010	Suess FB	45.6	0.2	3.60	1	pM	0	5.2	144	
27028	Suess D	46.5	4.6	6.88	1	pM	0	12.8	144	Cool halo
27037	Suess	47.7	4.4	9.14	1	pM	0	32.4	150	Cool halo, Radar "G"
27039	Suess B	47.4 W	5.6 N	8.24	1	pM	0	12.9	150	Cool halo
30043 X	Flammarion A	2.5 W	2.1 S	—	—	—	—	9.1	108	
30049 X	Floor (Herschel)	2.4	5.7	—	—	—	—	10.2	108	
30058 N	Near (Herschel C)	3.50	5.00	(10.41	1	C	0)	4.0	108	
30076 N	Near (Herschel B)	4.6 W	3.8 S	(6.33	1	pMC	0)	5.3	108	

30095	Mösting A	5.2 W	3.2 S	13.05	1	C	0	34.3	108	
31001 R	Rim (Mösting)	5.83	0.28	(26.49	2	pM	p)	13.9	108	
31019	Lalande C	6.9	5.6	10.64	1	C	0	7.0	108	
31043	Mösting C	8.1	1.8	3.84	1	pM	0	30.1	113	
31047	Lalande	8.6	4.5	24.13	1	C	p	30.9	113	
31081 N	Near (Turner B)	10.8	1.0	(4.73	1	pM	0)	16.4	113	
31143 X	No Feature	8.6	7.7	—	—	—	—	7.9	113	
31161	Lalande A	9.8	6.6	13.24	1	pMC	0	16.4	113	
32021 N	Near (Turner)	13.2	1.0	(11.85	1	pM	0)	8.9	113, 120	
32029 X	On (Lalande η)	13.1	5.2	—	—	—	—	6.0	113	Very rough
32067	Fra Mauro H or HA	15.5	4.1	6.26	1	pMC	0	12.8	120	
32083	Fra Mauro G	16.3	2.2	5.87	1	C	0	6.5	120	
32122	Bonpland F	19.3	7.3	3.98	1	pMC	0	9.3	120	
32153	Parry F	14.7	7.6	3.83	1	pM	0	11.5	120	
32199 X	Floor of (Bonpland)	17.1	7.9	(59.81	4	aMC	0)	7.3	120	Intersection of 2 rilles
33008	Fra Mauro D	17.6	4.8	4.64	1	pMC	0	14.7	120	
33056 X	No Feature	20.6	3.7	—	—	—	—	6.+	120	
33059	Fra Mauro A	21.0	5.5	9.87	1	pM	0	4.+	120, 125	Cool halo
33069	Fra Mauro C	21.6	5.5	6.71	1	pM	0	10.4	125	
33081 X	No Feature	22.5	0.9	—	—	—	—	4.0	125	
33083 X	No Feature	22.8	2.0	—	—	—	—	4.0	125	
34006 X	No Feature	23.9	3.6	—	—	—	—	6.3	125	
34030 R	Rim (Lansberg)	26.9	0.0	—	—	—	—	6.+	125	Rim
34048 R	Rim 1 (Euclides P)	26.7	4.8	—	—	—	—	3.5	125	
34069 R	Rim 2 (Euclides P)	27.9	5.6	—	—	—	—	4.9	132	Rim
34073 N	Near (Lansberg B)	28.1	2.2	—	—	—	—	7.6	132	
34182	Euclides	29.5	7.4	13.07	1	pM	0	11.2	132	
35003	Lansberg E	30.3	1.8	5.53	1	pM	0	10.5	132	
35005	Lansberg D	30.6	3.0	11.35	1	pM	0	27.5	132	
35007 X	No Feature	30.2 W	4.6 S	—	—	—	—	6.9	132	

(continued)

TABLE III (*continued*)

L & PL Number	Designation	Coordinates λ long.	β lat.	Diam (km)	C	B	CE	ΔT (K)	LO IV	Remarks
35078 X	No Feature	35.5 W	4.8 S	—	—	—	—	13.0	137	Moderately smooth
35093 X	No Feature	36.4	2.2	—	—	—	—	6.8	137	
36008	Wichmann C	37.4	4.7	2.75	1	pM	0	8.5	137	
36023 X	No Feature	38.6	2.1	—	—	—	—	4.5	137	Cool halo
36057 X	No Feature	41.0	4.3	—	—	—	—	8.7	143	
36097	Flamsteed	44.3	4.5	20.61	2	pM	p	27.5	143	
36113 N	Near (Wichmann)	38.6	7.6	(10.64	1	pM	0)	8.1	137	
36132 N	Near (Wichmann B)	39.6	7.1	(3.91	1	pM	0)	8.2	137	
36173	Flamsteed A	42.9	7.9	12.71	1	pM	0	27.2	143	
36180	Flamsteed B	43.7	4.9	9.54	1	pM	0	30.4	143	
37003 X	No Name Crater	45.2	1.8	~3.0	1 ?	—	—	7.8	143	
37016 X	No Feature	45.8	3.8	—	—	—	—	9.2	149	Moderately rough
37023 X	No Feature	46.6	2.2	—	—	—	—	7.1	149	
37029 X	No Feature	46.7	4.4	—	—	—	—	7.0	149	
37052 X	No Feature	49.0	1.2	—	—	—	—	6.2	149	
37058 X	No Feature	49.6	5.1	—	—	—	—	7.2	149	Rough
37101 X	No Feature	45.5	6.8	—	—	—	—	7.0	149	Moderately rough
37113 X	No Feature	46.4 W	7.8 S	—	—	—	—	6.7	149	Smooth
40007	Reaumur A	0.2 E	4.3 S	14.70	2f	C	0	4.9	101	
40040	Reaumur D	2.8	0.2	4.21	1	pMC	0	4.0	101	
40053	Seeliger	3.0	2.2	8.52	1	pMC	0	5.5	101	
40132	Hipparchus K	2.2	6.9	11.65	1	C	0	4.8	101	
40180	Hipparchus X	5.0	5.8	17.35	4f	C	0	5.3	096	
41006	Horrocks	5.9	4.0	30.50	2	pM	pp	6.4	096	
41025	Pickering	7.0	2.9	15.99	1	C	0	21.9	096	
41028	Hipparchus G	7.4 E	5.0 S	14.53	1	C	0	13.7	096	

41031 X	No Feature	8.0 E	0.6 S	—	—	—	—	4.+	096	Rough
41087 N	Near (Saunder T)	10.5	4.1	(6.36	1	C	0)	5.1	096	
41142	Hipparchus C	8.2	7.4	16.83	1	C	0	13.8	096	
42006 N	Near (Saunder A)	11.9	3.8	(7.16	1	C	0)	10.7	089, 096	
42010 X	No Name Feature	12.4	0.1	—	—	—	—	6.7	089	Moderately rough
42060	Theon Senior A	15.4	0.2	5.28	1	C	0	7.0	089	
43029	Alfraganus	19.0	5.4	20.79	1	C	pp	16.8	089	
43044 X	No Name Crater	20.1	2.8	3.5	—	—	—	13.0	084	Very rough
43058 X	No Feature	20.7	5.0	—	—	—	—	8.6	084	
44001	Moltke	27.7	0.6	6.47	1	pM	0	20.2	084, 077	
44025	Torricilli H	25.3	3.3	7.21	1	pM	0	5.7	077	
44044 N	Near (Torricelli C)	26.2	2.4	(11.09	1	pM	0)	6.5	077	
44067	Torricelli T	27.5	4.2	3.35	1	pM	0	5.6	077	
44081	Censorinus K	28.8	1.0	4.07	2	pM	0	5.9	077	
44084	Torricelli B	29.2	2.6	6.92	1	pM	0	30.4	077	
44087 X	No Name Feature	29.0	4.4	—	—	pMC?	—	12.9	077	Rough
45030	Censorinus (Near ?)	32.7	0.4	3.82	1	C	0	25.5	073, 072	Radar "M"
45056 R	Rim (Isidorus D)	34.6	4.0	(15.33	1	C	0)	11.5	072	Rim
45143 N	Near (Isidorus A)	33.2	8.0	(11.73	1	C	0)	10.2	072, 077	
45180 X	No Feature	36.0	6.0	—	—	—	—	8.8	072	Radar "N"
46061 X	No Feature	41.9	0.7	—	—	—	—	8.4	065	Moderately rough
46080 N	Near (Secchi)	43.5	0.5	(5.63	1	pM	0)	5.3	065	
46101	Cappela D	37.5	6.7	8.05	1	C	0	9.2	072	
47010	Secchi K	45.5	0.2	6.41	1	pM	0	8.3	060	
47023	Messier A	46.9	2.0	13.00	1	pM	0	46.0	060	Radar "P"
47026	Messier D	46.3	3.6	7.93	1	pM	0	17.6	060	
47041	Messier B	48.0	0.9	6.36	1	pM	0	9.6	060	
47150 X	No Feature	49.5 E	5.9 S	—	—	—	—	9.7	060	Cool halo, smooth

Planetary Laboratory Catalog of the University of Arizona (Arthur *et al.*, 1963, 1964, 1965, 1966). In the case of anomalies for which no identifiable or cataloged feature could be found on the AIC chart (designated by "X"), the L & PL number was calculated from its selenographic position. "R" indicates the anomaly is found on the rim of a crater, and "N" indicates it is found near a cataloged feature. In columns 3–8 information from the L & PL catalog relating to the position, size, class, background, and central peak descriptions are listed. Table IV explains the nomenclature used for this information. Column 9

TABLE IV

L & PL NOMENCLATURE REGARDING CRATER DESCRIPTION

Code	Symbol	Description
C — Crater classification		
	1	Complete and sharply defined rims
	2	Rims blurred
	3	Rims broken
	4	Ruins
	5	Battered and fragmentary
	f	Flooded interior
B — Crater background		
	M	Mare
	C	Upland
	MC	Overlap mare and upland
	a	Pre-mare (for M and MC only)
	p	Post-mare (for M and MC only)
CE — Central peak information		
	0	No central object
	P	Single conspicuous central peak
	PP	Multiple conspicuous central peaks
	p	Weak central peak
	pp	Multiple weak central peaks
	R	Central ridge

lists the measured temperature differential ΔT of each feature over its environs. Column 10 lists the Lunar Orbiter IV photograph frame number where the anomaly is located.

Some explanation is necessary regarding the limitations inherent in the contours. Besides the position errors already discussed, there is a certain amount of noise present in the system which comes from the sky, the detector, and the electronics. This was determined by measuring the rms noise present when the detector was off the Moon and scanning the sky. At a brightness temperature of 150 K, this noise was equivalent to ± 4 K, at 175 K to ± 1.5 K, and at 200 K to ± 0.8 K. Further, as previously described, if the thermal feature is comparable to the size of the resolution element, the signal detected is very sensitive to its

position in the scan raster pattern. In this regard, it has been possible to infer the size of the thermal feature from the number of data points along a scan line which correspond to observation of the anomaly. In the chart descriptions frequent reference is made to this procedure which took into account the effect of diffraction in enlarging the apparent size of the anomaly. In a number of cases an areal correction has been made to the observed ΔT to take into account dilution of the anomalous signal by emission from the environs.

Finally, radar images of the equatorial band of the Moon have been reported (Massachusetts Institute of Technology Lincoln Laboratory Report, 1968). These measurements, made at 3.8 cm, show some remarkable correlations with this infrared data. They are particularly interesting since the resolution attained was substantially better than that for the infrared results. In the AIC chart descriptions, references will be made to these radar data; in particular, eighteen specific radar features were identified (by letter) in the report. When they correspond to thermal anomalies, this will be noted in the remarks column of Table III. It is to be noted that no radar features were listed closer than 23° from the disk center for reasons described in the above report.

VI

DESCRIPTIONS OF THE THERMAL OVERLAYS

Lunar Orbiter IV photographs for the most prominent thermal anomalies were studied and are referred to in the descriptions that follow. It is interesting to note the variety in the surface features which are thermally anomalous. It is equally interesting to note that other surface features, although very similar in appearance, are *not* anomalous.

A. *Maestlin AIC 57 D (40°–50°W, 0°–8°N)*

This chart covers a region of Oceanus Procellarum southwest of Kepler. Examination of the contours shows that apparently on this chart there was a systematic misregister in position, since a number of the anomalies maximize a fraction of a degree west in longitude from their probable source of emission.

Three anomalies are associated with diffuse 3.8-cm radar enhancements. For instance, the anomaly associated with Suess (27037) has contours which are elongated in the southwest direction toward Suess G. Examination of the radar image shows a diffuse enhancement between these two craters at 48.0°W, 3.9°N. It appears that this feature and Suess are both thermal anomalies and are so close together that they are not resolved in the infrared data. The Lunar Orbiter IV Frame 150 photograph shows a sharp crater 1.6 km in diameter between Suess and Suess G; it is surrounded by a bright region about 7 km in extent.

A radar enhancement was reported at 42.3°W, 7.0°N which corresponds to the thermal anomaly (26172 X) with $\Delta T = 18.2$ K. This feature in a ray extending westerly from Kepler is near a crater 3 km in diameter; the LO IV Frame 144 photograph shows this crater to be surrounded by numerous smaller craters.

Encke X (26041), with a $\Delta T = 17.1$ K, is an anomaly associated with a diffuse radar enhancement at 40.3°W, 1°N. The LO IV Frame 144 photograph shows Encke X to be a sharp crater surrounded by a bright ray system. There is a cluster of smaller craters to the southwest.

Thermal anomalies are observed on Suess B (27039), Maestlin H (26088), and Suess F (27002); the radar images show these craters to be bright, with little if any enhancement exterior to the craters. Several cool areas are in this chart, for example, at 42.9°W, 1.9°N and 40.2°W, 6.7°N.

B. *Encke AIC 57 C (30°–40°W, 0°–8°N)*

The southern portion of the prominent ray crater Kepler (26104) is on this chart. Although not shown here, the signal maximizes on the northern interior rim of the crater with $\Delta T = 27.1$ K. The nearby crater Kepler A (25182) has a $\Delta T = 23.2$ K, almost as large as Kepler itself; if, as suggested by the extent of the contours in the scan direction, the anomalous emission comes only from the crater, a corrected $\Delta T = 51$ K is calculated. This may indicate Kepler A is a (thermally) younger crater than Kepler.

More subdued enhancements are observed on Hortensius A (25007), Lansberg A (25010), and Encke B (25094). An anomaly (26048 X) is present at 40°W, 4.7°N where no radar feature is noticeable; the LO IV Frame 144 photograph shows a bright crater about 4 km in diameter.

C. *Reinhold AIC 58 D (20°–30°W, 0°–8°N)*

The region covered by this chart includes the southwestern portion of the Copernican ray system. The strongest anomaly with $\Delta T = 27.7$ K is observed on Hortensius C (24140), a small bright crater 6.74 km in diameter. If the emission is assumed to come from the crater itself, a corrected $\Delta T = 94$ K is calculated; 32% of the interior of the crater being bare rock with $\gamma = 20$ would account for this temperature differential. Hortensius (24161), 14.65 km in diameter, shows a $\Delta T = 11.3$ K.

Reinhold (23085 R) shows a subdued anomaly on the southwestern rim. Two anomalies in this chart seem to be associated with diffuse radar enhancements. The first identified as Reinhold NA (24022) is at 25.5°W, 2°N. This crater has a diameter of only 2.24 km and is entirely too small to account for the observed $\Delta T = 20$ K. The 3.8-cm radar image shows a diffuse feature at this position about 15 km in extent. The second anomaly (24018 X) at 24.3°W,

4.8°N has a $\Delta T = 5$ K; it appears as a diffuse radar enhancement. A full-moon photograph shows this to be a bright region with a small crater at its center.

D. *Gambart AIC 58 C (10°–20°W, 0°–8°N)*

Four well-defined anomalies, all associated with craters, are present in this chart. Gambart C (22005) is 12.18 km in diameter and has a $\Delta T = 35.0$ K. The extent of the contours suggest the source of emission is probably confined to the crater; this thermal feature was photographed by LO II and may be found on Frame 112 (see Fig. 31).

Gambart B (22003) and Gambart G to the southwest apparently are two anomalies which merge together in the contours because of their close proximity to each other. Gambart A (23021) with $\Delta T = 29.7$ K is a crater 11.98 km in diameter. The extent of the contours suggest the emission comes from the crater
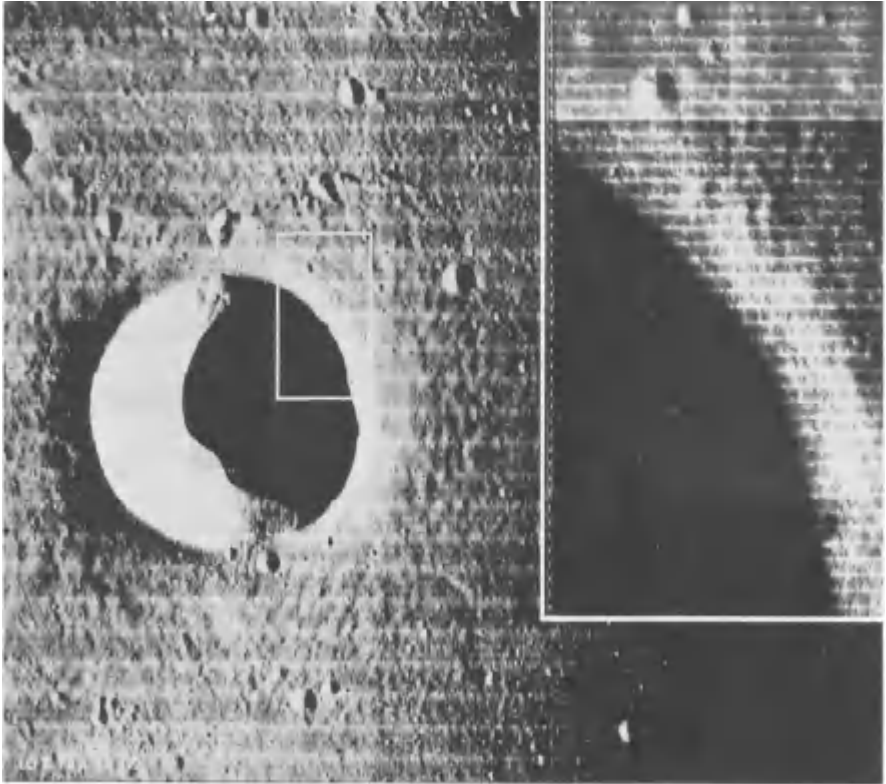
FIG. 31. Photograph of Gambart C (22005) taken by Lunar Orbiter II. This crater has been identified as a hot spot with $\Delta T = 35$ K. The small inset is a portion from the high resolution camera.

itself; with this assumption, an areal correction can be calculated yielding a corrected $\Delta T = 52$ K.

Copernicus H (23112) is a small sharp dark halo crater in the ray system of Copernicus. The extent of the contours indicates that the source of the emission probably extends beyond the crater and may correspond to the area of the dark halo.

E. *Pallas AIC 59 D (0°–10°W, 0°–8°N)*

This chart covers upland regions and a portion of Sinus Medii. There are relatively few anomalies on this chart. Bode (20141) has a rather mild anomaly with $\Delta T = 10.2$ K apparently centered outside the west rim. Examination of a full-moon photograph reveals a diffuse white patch just exterior to the crater. On the LO IV Frame 109 photograph this patch is shown to be centered on a sharp crater about 1.6 km in diameter. Since the extent of the contours in the scan direction indicate the source of emission is quite small, it is probable that the anomaly is associated with this crater.

Two other craters, Bode G (20161) and Schröter D (21067) are anomalies. A thermal feature (21042 X) at 8.1°W, 1.5°N is probably associated with a bright crater at 7.8°W, 1.6°N apparent in the LO IV Frame 114 photograph. The Surveyor VI landing site is on this chart.

F. *Triesnecker AIC 59 C (0°–10°E, 0°–8°N)*

This chart covers a portion of Sinus Medii and upland regions. Triesnecker (10067) is a ray crater 25.95 km in diameter with a $\Delta T = 17.9$ K. The extent of the contours in the scan direction indicate the source of infrared emission is probably confined to the crater. Bruce (10002), Blagg (10022), and Chladni (10016) are all anomalies. Rhaeticus B (11012) is interesting because it is an anomaly while several similar craters nearby (e.g., Rhaeticus G and D) are not. Examination of LO IV Frame 97 reveals that Rhaeticus B has a sharper rim and hummocky terrain exterior to the crater, suggesting it is probably a younger feature than the others. Similarly, Triesnecker D (11006) is anomalous while several similar craters nearby show no enhancement.

An anomaly associated with Pallas N (10102) may possibly be associated with a bright patch just to the northwest of the crater.

The Hyginus Rille has a thermal enhancement extending on either side of the crater Hyginus (11103). In this AIC chart a thermal maximum is shown about 10 km east of this crater. On previous scans another maximum is shown in the approximate position of Hyginus; apparently this anomaly is rather small in extent since its detection is critically dependent on the position of the scan raster.

G. *Agrippa AIC 60 D (10°–20°W, 0°–8°N)*

The most prominent anomaly in this chart is the bright ray crater Dionysius (12094), which is 17.64 km in diameter with $\Delta T = 49.3$ K. The extent of the contours in the scan direction suggest the anomaly is probably confined to the crater, although the region of high albedo extends about a crater diameter beyond the rim. LO V photography shows this crater has rough interior slopes with a festooned floor.

Godin (11073), which has a bright rim and a ray system apparent at full moon, has a $\Delta T = 33$ K. The nearby crater Agrippa (11097 R), which is somewhat larger but similar in appearance (except for the rays), has an anomalous region on the interior east rim with $\Delta T = 10.5$ K.

Silberschlag (12110), a bright crater 13.42 km in diameter, has a $\Delta T = 7$ K whereas the nearby Silberschlag A, a similar appearing crater 6.88 km in diameter, has $\Delta T = 27.3$ K. Rima Ariadaeus passes between these two craters; to the west at 11.3°E, 7.3°N is a subdued elongated anomaly which may be associated with the rille.

Sosigenes A (13113), a class 1 crater 11.87 km in diameter, has a rim only slightly brighter than the surrounding mare at full moon and has a $\Delta T = 16.4$ K. Subdued anomalies are observed on Whewell (12037), Cayley (12066), Ritter B (13025), and Schmidt. A relatively flat area (12041 X) at 14.2°W, 1°N is an anomaly with $\Delta T = 7$ K; no major feature can be identified in the LO IV photography that could account for the enhancement.

H. *Arago AIC 60 C (20°–30°E, 0°–8°N)*

This chart covers the southwest portion of Mare Tranquillitatis and includes the Ranger VIII impact site and the Surveyor V landing site. The most prominent anomaly is Manners (13047), a class 1 crater 15.14 km in diameter with a $\Delta T = 29.1$ K. The extent of the thermal contours suggest the source of infrared emission is probably confined to the crater; this is consistent with the radar image which shows a bright interior with no enhancement exterior to the rim.

Arago (13160) and Arago D (13172) have essentially the same ΔT over their environs, although the former is 26.04 km diameter and the latter only 4.55 km. Thus the smaller (class 1) crater no doubt has a much higher corrected temperature than the larger (class 2) crater.

In this chart almost every crater larger than several km in diameter appears to be anomalous. An exception is Sabine (20°E, 1.4°N), a class 2 crater 30.3 km in diameter, which is thermally bland. There are a number of wrinkle ridges in the mare which show no apparent thermal structure. Maskelyne G (14044) is an example of one of the so-called infrared cool halo craters. The first manned landing (Apollo 11) is located in a thermally bland region, 23°.43 E, 0°.69 N.

I. *Maskelyne AIC 61 D (30°–40°E, 0°–8°N)*

This chart covers the southeast portion of Mare Tranquillitatis. The most prominent anomaly in this region is Maskelyne (15003) with a $\Delta T = 17.6$ K. The maximum signal is obtained on the east rim, which at full moon appears to be the brightest portion of the crater.

Lesser anomalies are observed on Maskelyne H, Cauchy M, and Cauchy CA. The latter crater, less than 4 km in diameter, has a $\Delta T = 8.4$ K. If it is assumed the anomalous emission is confined to the crater alone, calculations show that the crater temperature must be 256.7 K, yielding a corrected $\Delta T = 82.5$ K. If it is further assumed that the interior of the crater consists of soil (at a temperature of the environs) and bare rock, one can infer that 27% of the interior must be exposed rock. It is not possible from the LO IV Frame 66 photograph to confirm this speculation.

Cauchy ω, a dome with a summit crater at 37.7°E, 7.2°N, appears to be slightly cooler than its environs.

J. *Secchi AIC 61 C (40°–50°E, 0°–8°N)*

The northwest portion of Mare Fecunditatis is covered in this chart. Four anomalies, Taruntius E, F, G, and H are craters similar in size and albedo. A mild anomaly is noted on Taruntius C (17110), a sharp rimmed crater on the rim of Taruntius. Taruntius B and TB both appear to be anomalous, but because of the position of the scan raster appear as a single thermal feature.

Secchi B (41.6°W, 3.75°N) shows a thermal anomaly, whereas a radar anomaly is reported at 41.4°W, 3.1°N, which is near Secchi A. Close examination of the radar image suggests that both of these craters are resolvable radar enhancements; they are, however, not resolvable in the infrared contours.

K. *Flamsteed AIC 75 A (40°–50°W, 0°–8°S)*

The major structural feature in this chart is the Flamsteed ring in which Surveyor I landed. The three major anomalies are Flamsteed (36097), Flamsteed B (36180), and Flamsteed A (36173). The thermal maxima for these features are all displaced ~0.3° to the northwest from the crater centers, which is probably caused by a slight error in the grid fit in this region. There is a small thermal low just to the southeast of Flamsteed. It should be noted that often such a low is found next to a sharp anomaly; it can be instrumental, caused by overshoot in the electronic circuitry (in the scan direction).

Three subdued anomalies 37003 X, 37023 X, and 37052 X appear to be associated with 3.8-cm radar enhancements. Two others, 37016 X and 37101 X, do not appear to be radar features.

In this chart there are two cool regions (40.5°W, 1°S and 49.8°W, 5.75°S) associated with rough areas covered by multiple craters.

L. *Wichmann AIC 75 B (30°–40°W, 0°–8°S)*

There are relatively few thermal anomalies in this region of Oceanus Procellarum. The most prominent is Lansberg D (35005) with $\Delta T = 27.5$ K. The extent of the contours in the scan direction suggest the anomaly is confined to or within the crater. Lesser anomalies are observed on Lansberg E, Wichmann, and Wichmann B. In this chart the thermal maxima for anomalies associated with craters all seem to be displaced somewhat to the west.

A mild anomaly (35078 X) at 35.5°W, 4.8°S is not associated with a crater. It appears as a slight enhancement in the 3.8-cm radar image; the LO IV Frame 137 photograph shows a number of km-sized craters in this region.

M. *Euclides P AIC 76 A (20°–30°W, 0°–8°S)*

There are not many anomalies in this chart, which includes both upland and mare regions. The crater Euclides (34182) has the largest ΔT of 11 K; although the contours appear elongated it is probably due to the circumstance that the anomaly happened to be centered between adjacent scan lines so that the contours are stretched in the direction perpendicular to the rapid scan. A subdued anomaly is present on the northwest interior rim of Lansberg. A cool region at 25.5°W, 2°S is apparently associated with a rough mountainous region. The Surveyor III landing site at 23.17°W, 3.33°S and the Apollo 12 landing site, Ocean of Storms, at 23°.30 W, 2°.40 S are in thermally bland regions.

N. *Fra Mauro AIC 76 B (10°–20°W, 0°–8°S)*

The more prominent anomalies in this chart are associated with craters, including Turner B, Turner, Fra Mauro H and HA, and Fra Mauro D.

Lesser anomalies are present on other features, such as Lalande η (32029) where the maximum signal was observed on the southern portion of this ridge. The enhancement observed on Parry F (32153) is elongated in a southwest–northeast direction along Rima Parry. The mild anomaly (32199 X) on the floor of Bonpland appears to be associated with the intersection of two rilles. The Apollo 14 site, Fra Mauro, is located at 17°.46 W, 3°.66 S in a thermally bland region.

O. *Flammarion AIC 77 A (0°–10°W, 0°–8°S)*

This chart covers mostly upland regions except for the southwest portion of Sinus Medii. It includes the small crater Mösting C (31043) with a $\Delta T =$

30.1 K. Preliminary calculations suggested that if the anomalous region were confined to the crater itself (3.84 km diameter), then the calculated temperature for the crater would be consistent with its surface being bare rock with $\gamma = 20$ (Shorthill and Saari, 1965). More precise calculations now indicate an even higher areally corrected temperature for the crater, implying a value of γ less than 20. Since this is unlikely, we conclude that the source of anomalous emission must come from a region larger than the crater itself. This conclusion is supported by the high resolution photographs of this crater obtained from LO V, which reveal a hummocky terrain extending about a crater diameter beyond the rim. In this region, a large number of rocks are evident.

The extent of the contours over Mösting A (30095) suggest the anomaly ($\Delta T = 34.3$ K) is confined to the crater, or within 1 km of the rim. In the LO IV Frame 108 photograph hummocky terrain surrounding the crater is not particularly evident.

The thermal anomaly shown on Lalande (31047) is elongated in the north–south direction. A subdued anomaly on Herschel is centered on the western portion of the crater. Similarly, the anomaly on Mösting is positioned on the northern interior rim.

South of Palisa U is a lesser anomaly (31143 X) at 8.6°W, 7.7°S without an identifiable feature. The LO IV Frame 113 photograph shows a sharp crater ~1 km in diameter with two dozen smaller craters clustered around it.

P. *Hipparchus AIC 77 B (0°–10°E, 0°–8°S)*

In this upland region the three most prominent anomalies, Pickering (41025), Hipparchus G (41028), and Hipparchus C (41142) are almost the same size and appear as bright spots on the full moon. The LO IV Frame 96 photograph shows them to be sharp craters with flat floors. Pickering has, however, a $\Delta T = 21.9$ K, half again as much as the 13.7 K and 13.8 K values observed on the other two features. A number of rather subdued enhancements are present on other craters in this chart.

Q. *Delambre AIC 78 A (10°–20°E, 0°–8°S)*

This chart covers an upland area southwest of Mare Tranquillitatis. Alfraganus (43029) is the most prominent anomaly with a $\Delta T = 16.8$ K. Rather subdued enhancements are observed on Theon Senior, Theon Senior A and C, and Theon Junior. An anomaly (42006 N) just northwest of Saunder A has a $\Delta T = 10.6$ K and is probably associated with a bright spot apparent on a full-moon photograph. The LO IV Frame 089 and 096 photographs do not reveal any particular feature in this region which would explain its being a thermal anomaly.

R. *Torricelli AIC 78 B (20°–30°E, 0°–8°S)*

The most prominent anomaly on this chart is Torricelli B (44084) with $\Delta T = 30.4$ K. This crater (6.92 km diameter) is surrounded by a region of dark albedo and has a small intense bright spot on the northeast rim. The extent of the thermal contours suggest the anomalous region is confined to the crater, in which case the corrected ΔT becomes 106 K. The radar image shows no apparent enhancement outside the crater. The thermal contours indicate an anomaly between Torricelli and Torricelli A; in another scan this feature is mapped as two separate anomalies because of a shift in the raster pattern.

Moltke (44001), slightly smaller than Torricelli B, has a $\Delta T = 20.2$ K. It is a bright crater surrounded by a region of high albedo extending about a crater radius beyond the rim. The radar image indicates an enhancement about as large, so it is likely that the thermally anomalous region is of the same extent.

A white patch at 20.7°E, 5°S is a "no feature" anomaly (43058 X) with $\Delta T = 8.6$ K. The anomaly is in a relatively flat upland region and contains two small craters. The extent of the contours suggest the anomaly is about the size of the bright patch. A similar anomaly (43044 X) is located at 20.1°E, 2.7°S on an unnamed crater ~3.5 km in diameter. The LO IV Frame 084 photograph reveals this crater is located on a ridge which is associated with Alfraganus A, a class 3 crater. A full moon photograph shows the crater to be in a white patch containing a number of bright dots.

S. *Capella AIC 79 A (30°–40°E, 0°–8°S)*

The most prominent anomaly on this chart is the ray crater Censorinus (45030). The extent of the thermal contours show that the anomaly must extend beyond the crater (which is only 3.82 km in diameter), and probably is associated with the larger area surrounding the crater which is characterized by high albedo and an enhancement in the 3.8-cm radar image.

Lesser thermal enhancements are observed on Isidorus, Isidorus E, Isidorus D, and Capella D. The anomaly (45180 X) at 36°E, 6°S falls between Capella C, CA, and RA and is a diffuse radar enhancement. The LO IV Frame 072 photograph shows two small craters (~1.5 km diameter) in this region. On a full-moon photograph the anomaly shows up as a bright spot about 10 km in diameter. The extent of the thermal contours indicate the anomaly is ~15 km in diameter.

T. *Messier AIC 79 B (40°–50°E, 0°–8°S)*

Most of this chart is covered by the western portion of Mare Fecunditatis and includes the unusual craters Messier and Messier A (47023). A strong infrared signal was received from these craters, with an indicated $\Delta T = 46.0$ K

over the environs. On this particular scan the anomaly appears to be maximum on the northeast rim of Messier A; another scan, during which these craters were intersected by the scan raster somewhat differently, indicates two distinct anomalies. Enhancements on these craters are indicated by the 3.8-cm radar images. The rays extending to the west from these craters show no thermal enhancement. The nearby craters Messier B and D are somewhat enhanced. Anomalies are observed at 41.9°W, 0.7°S and 49.5°W, 5.9°S although no sizeable craters are in the vicinity.

VII

DISCUSSION

The 165 thermal anomalies listed for the Apollo Band occur on a variety of features, although the majority are associated with craters. Ninety-eight of those anomalies were identified with craters listed in the L & PL catalog. Table V

TABLE V

CHARACTERISTICS OF 98 THERMAL ANOMALIES IN THE APOLLO BAND IDENTIFIABLE WITH CRATERS IN THE L & PL CATALOG

Crater class	Background			
	Mare	Boundary	Upland	Total
1	58	8	19	85
2	7	1	3	11
3	1	0	0	1
4	0	0	1	1
5	0	0	0	0
	66	9	23	98

shows the distribution of these craters with respect to their class and background. Eighty-five of these craters are class 1 and sixty-six are in maria. In fact, the table shows that in the Apollo Band over half of the anomalies associated with craters are class 1 and occur in maria.

Figure 32 shows the distribution of the 165 anomalies with respect to the observed ΔT over their environs. The fall-off at low values of ΔT is due, of course, to the criterion employed in designating a feature anomalous. The average ΔT for these anomalies is 12.3 K. Since many of the features are smaller than the sensor resolution and the signal is diluted with environ emission the distribution of the actual (corrected) ΔT's will be considerably different.
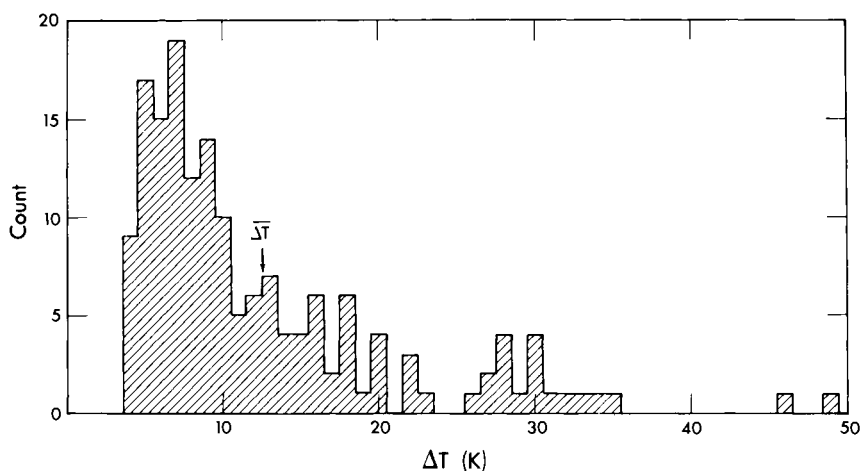
Fig. 32. Distribution of observed temperature difference ΔT above environs for the 165 identified thermal anomalies in the Apollo band. The average ΔT is 12.3 K.

Among the several possible explanations for the thermal anomalies, internal heating is particularly intriguing since it bears on whether the Moon is presently volcanically active. If it were possible to prove conclusively that some of the thermal anomalies were caused by internal heating, the implications regarding the history of the Moon would be far-reaching. In this regard, the correlation between the radar and thermal anomalies is interesting since the radar is sensitive only to the structure of the surface and not to its temperature. For example, a region which is strewn with rocks and/or characterized by exceptionally dense surface material will exhibit enhanced radar backscattering and anomalous eclipse cooling. On the other hand, if a region were not so characterized, but possessed a subsurface gradient sufficient to produce a thermal anomaly, it would present a bland appearance to a radar. It seems probable, therefore, that the most likely places on the Moon where internal heating may play a role are those thermal anomalies (if they exist) which are not radar anomalies. To this end, a critical comparison of the radar and infrared data would be useful. Actually, the search for internal heat sources could perhaps best be done by high resolution microwave radiometry from an orbiting spacecraft.

In this connection, the observation of lunar events, temporary changes in the appearance of lunar features (Middlehurst *et al.*, 1968), may be evidence of internal activity. The question arises whether there is any correlation between lunar events and thermal anomalies. While such a study has not been made, it is readily apparent that many sites of lunar events such as Alphonsus, Gassendi, Plato, and Sabine are not thermally anomalous. Others are, such as Tycho, Aristarchus, Messier A and Messier, and Theophilus. Thus a cursory examina-

tion of the data would suggest that there is no correlation between the two measurements.

It is possible that most thermal anomalies are produced by concentrations of rocks on the surface, in which case the infrared device is in reality a "rock detector." Certainly the Lunar Orbiter, Surveyor, Apollo, and Luna photographs show that rocks are a common feature of the lunar landscape, occurring in and around craters, on certain ridges, and in some of the rilles.

The so-called "no feature" anomalies are of particular interest. Many of these anomalies show a radar enhancement and, from the Lunar Orbiter photography, are often associated with a single, very small, bright crater or a cluster of craters, and in most instances with a region or "patch" of high albedo. Because of their unusual appearance, it is possible these features are an expression of a process occurring on the lunar surface which is essentially different from those which have formed other features, such as craters and rilles, which are thermally anomalous. High resolution photography, manned and automated exploration would be necessary to validate this speculation.

Note Added in Proof

Since the completion of the manuscript for this chapter, additional work bearing on the interpretation of the eclipse infrared observations has been performed. A comprehensive review of infrared observations of the moon was published by Shorthill (1970). At the Lunar Science Institute, Houston, Texas a comparison of infrared, radar, and geologic mapping of lunar features (Thompson *et al.*, 1970) was conducted. Fifty-two craters, the strongest of the 1000 thermal anomalies, were compared with 70 cm and 3.8 cm radar maps. Other representative groupings of features were also studied. The infrared and radar response of Apennine–Hadley Region (Apollo 15 site) was investigated (Zisk *et al.*, 1971).

References

Arthur, D. W. G., Agnieray, A. P., Horvath, R. A., Wood, C. A., and Chapman, C. R. (1963). *Commun. Lunar Planet. Lab.* **2**, No. 30.

Arthur, D. W. G., Agnieray, A. P., Horvath, R. A., Wood, C. A., and Chapman, C. R. (1964). *Commun. Lunar Planet. Lab.* **3**, No. 40.

Arthur, D. W. G., Agnieray, A. P., Pellicori, R. H., Wood, C. A., and Weller, T. (1965). *Commun. Lunar Planet. Lab.* **3**, No. 50.

Arthur, D. W. G., Pellicori, R. H., and Wood, C. A. (1966). *Commun. Lunar Planet. Lab.* **5**, No. 70.

Massachusetts Institute of Technology Lincoln Laboratory Report. (1968). *Radar Stud. Moon* **2**, 1.

Middlehurst, B., Burley, J. M., Moore, P., and Welther, B. L. (1968). *NASA Tech. Rep.* **TR R-277**.

Saari, J. M., and Shorthill, R. W. (1965). *Nature (London)* **205**, 964.
Saari, J. M., and Shorthill, R. W. (1967). *NASA Contract. Rep.* **NASA CR-855.**
Saari, J. M., Shorthill, R. W., and Deaton, T. K. (1966). *Icarus* **5**, 635.
Samaha, A. H. (1967). "Measure of the Moon," p. 78. Dordrecht.
Shorthill, R. W. (1970). *J. Spacecr. Rockets* **6**, No. 4, 385.
Shorthill, R. W., and Saari, J. M. (1965). *Science* **150**, 210.
Thompson, T. W., Masursky, H., Shorthill, R. W., Zisk, S. H., and Tyler, G. L. (1970). Contribution No. 16, Lunar Science Institute, Houston, Texas.
Zisk, S. H., Carr, M. H., Masursky, H., Shorthill, R. W., and Thompson, T. W. (1971). *Science* **173**, 808.

Infrared Emission from the Surface of the Moon

DONALD F. WINTER*
Boeing Scientific Research Laboratories
Seattle, Washington

I

INTRODUCTION

The transfer of heat energy in the lunar crust has been the subject of intensive inquiry for many years. It was long ago recognized that radiation from the Moon at micron wavelengths could hold important clues to certain properties of its surface. The first temperature measurements of the lunar surface were made nearly a century ago, although the earliest work of real significance was performed by Pettit and Nicholson of Mount Wilson Observatory (Pettit and Nicholson, 1930; Pettit, 1940). These investigators reported a variety of results, including subsolar point temperatures of 407°K and 358°K at full moon and quarter phase, respectively, and an antisolar point temperature of 120°K. They also measured the radiative intensity across the full-moon disk and the transient temperature of an area near the center of the disk and near the limbs during an eclipse.

Although the problem of interpreting lunar surface temperatures had been considered from a theoretical standpoint prior to 1930 (Epstein, 1929), the data acquired by Pettit and Nicholson inspired investigators to perform more thorough analyses. In particular, these observations made it possible to infer thermal properties and structural characteristics of the surface by comparing measured temperature variations with predictions from hypothetical models based on various assumptions regarding thermal energy transfer. It was assumed

* Present address: Department of Oceanography, University of Washington, Seattle, Washington.

in such studies that the brightness temperatures reported could be regarded as true surface temperatures. The first definitive theoretical treatment was that of Wesselink (1948) who compared reported temperature variations with calculations based upon a model which characterized the lunar surface as a semi-infinite continuous medium with uniform mass density ρ, specific heat c, and thermal conductivity k. According to this model, a useful measure of the transient temperature response of the medium is the magnitude of the thermal parameter $\gamma = (k\rho c)^{-1/2}$. Comparing his theoretical results with the eclipse data of Pettit and Nicholson, Wesselink found that the inferred value of γ was of the order of one thousand. With reasonable values of ρ and c, this implied that the thermal conductivity of the lunar surface material was lower than that of terrestrial rock by a factor of a thousand. The conclusion of the study was that the surface of the Moon was probably porous or covered with dust.

Subsequent studies by Jaeger and Harper (1950) and Jaeger (1953) generally supported this view although they achieved somewhat better agreement with the Pettit and Nicholson data by using a two-layer model with a layer of insulating dust overlying a homogeneous rock substrate. The uniform properties model was subsequently reexamined by Krotikov and Shchuko (1963) using a computer to perform the numerical integrations of the heat equation. The variation of surface temperature over a lunation was obtained with greater precision than in the earlier work and the conclusions regarding the porous or particulate nature of the lunar surface were further substantiated.

Calculated lunar surface temperatures could be compared with brightness temperatures inferred from measured radiative intensities because the optical depth of the lunar soil at micron wavelengths is small compared with the penetration depth of temperature variations. However, since the temperature profile within the lunar crust undergoes significant excursions only to a depth of the order of a decimeter, infrared measurements provide information pertaining only to the uppermost regions of the regolith. On the other hand, contributions to radio emission from the Moon's surface at millimeter wavelengths originate from depths of the order of a meter. Theoretical calculations of the apparent radio brightness of the Moon traditionally assumed that the temperature variation within the lunar soil could be specified independently from a suitable thermophysical soil model. Expressions for the temperature–depth profile, which proved particularly useful in this regard, were derived from the surface temperature appropriate to constant properties models, such as those described above. However, attempts to explain both the infrared and microwave observations on the basis of these assumptions led to certain inconsistencies. For example, although the general features of the infrared data could be reproduced by a surface material characterized by a thermal parameter γ of the order of one thousand, the microwave data were more satisfactorily explained by a thermal parameter of the order of 350. It was

proposed that the infrared and microwave measurements might be reconciled by postulating a surface layer, extending at least to a depth of decimeters, which exhibited temperature-dependent thermal properties and which behaved like a good insulator during the lunar nighttime (Muncey, 1958, 1963; Krotikov and Troitsky, 1963; Linsky, 1966). In any event, the data in both wavelength regimes were suggestive of highly porous or fine-grained material whose thermal properties depend upon temperature and, possibly, upon depth.

These general inferences drawn from infrared data regarding the nature of the lunar soil have been verified by recent *in situ* observations. The regolith at the landing sites of Luna, Surveyor, and Apollo spacecraft appears to be comprised of a fine, moderately cohesive, graded soil with a particle size distribution which extends to at least a few tens of microns. Although it would be premature to conclude that the entire lunar surface has properties identical with the soil found in the vicinity of the landing sites, these areas are probably not unique. Even prior to instrumented landings, however, several independent lines of evidence indicated that much of the lunar surface was overlain with a fine-grained soil whose porosity, particle size, and cohesiveness might vary with distance below the surface. It also became apparent that thermophysical soil models would have to be refined to interpret new data acquired by increasingly sophisticated techniques. Several investigators reexamined the problem of heat transfer in the lunar crust and included several effects which had been neglected in earlier treatments. Before describing these efforts, we shall briefly review some of the more recent experimental measurements at infrared wavelengths which have proved useful in evaluating various theoretical models and, more generally, in specifying the radiation environment on the lunar surface.

II

RECENT INFRARED OBSERVATIONS

The apparent temperature of a region of the lunar surface is inferred from the measured intensity of emitted radiation over a frequency range which is determined by the wavelength characteristics of the filters used with the apparatus. The inference would be fairly straightforward in the case of a total energy detector if the region resolved by the detector were isothermal. On the other hand, in the case of a narrow band detector with center wavelength λ, the received intensity over a small wavelength interval is weighted by a factor of the form $1/[\exp(T_c/T) - 1]$, where $T_c = hc/\lambda K$ is the characteristic temperature of the detector. The ratio of the detector response to a surface at temperature T to a surface at temperature T_0 is given by the response function

$$R(T) = \frac{\exp(T_c/T_0) - 1}{\exp(T_c/T) - 1}\,. \tag{1}$$

As pointed out by Hopfield (1967) and Roelof (1968), if the region of the lunar surface within view is not isothermal, then the contributions to the total signal from areas or objects at various temperatures are weighted differently by the detector. Therefore, the apparent temperature ascribed to the region can, under some circumstances, depend rather critically upon the response function.

Another difficulty in the interpretation of intensity measurements over the lunar disk arises from the fact that the emitting surface is not locally smooth. In fact, the Moon's surface is fairly rough on a scale of centimeters, due to a proliferation of small craters and surficial rock. As a consequence of local roughness, the directional emissive properties of an area at a given location on the lunar surface will depend in a nontrivial way on the local sun elevation angle and the angle of observation. Temperatures are generally inferred from intensity profiles of scans across the disk made at a variety of lunar phase angles, and therefore include data acquired over a wide range of aspect angles under differing conditions of insolation. Most reported temperatures have not been modified by considerations of directional emission because the effect is not sufficiently well understood at present.

The foregoing remarks show that the deduction of a realistic lunar surface temperature from an infrared detector signal is not altogether a straightforward matter. In addition to the problems mentioned thus far, the reduction of the data presents rather special difficulties which have been discussed in considerable detail by Ingrao *et al.* (1966). Nevertheless, recent advances in experimental techniques have produced useful data which provide us with a fairly comprehensive picture of the Moon's surface as it appears at infrared wavelengths.

Figure 1 presents a collection of measurements reported by a number of observers, which gives an indication of the thermal history of an area on the lunar equator over the course of a lunation. In the figure, the apparent temperature T is plotted as a function of the fraction τ of the synodic period. The lunar daytime temperatures indicated by triangles in Fig. 1 were acquired by Sinton (1962) and coworkers at Lowell Observatory. Using a narrow-band filter at 8.8 μ, these observers constructed maps of isotherms of the illuminated surface at nine different phase angles. The lunar disk was scanned in a television-like raster using the earth's rotational motion and the motion of the Moon in declination. The resolution employed was about 25″ arc (1″ arc corresponds approximately to 1/1800 of a lunar diameter). The data in the figure were obtained from thermal contours at seven different lunar phase angles.

The second set of lunar daytime temperatures was provided by Saari and Shorthill (1967a) at Boeing Scientific Research Laboratories. These observers scanned the illuminated disk at a resolution of 8″ arc using a mercury-doped germanium detector cooled to liquid-hydrogen temperature and a bandpass filter admitting radiation in the wavelength interval 10 to 12 μ. Observations were made at 23 phase angles from $-125°$ to $+135°$ in both the infrared and
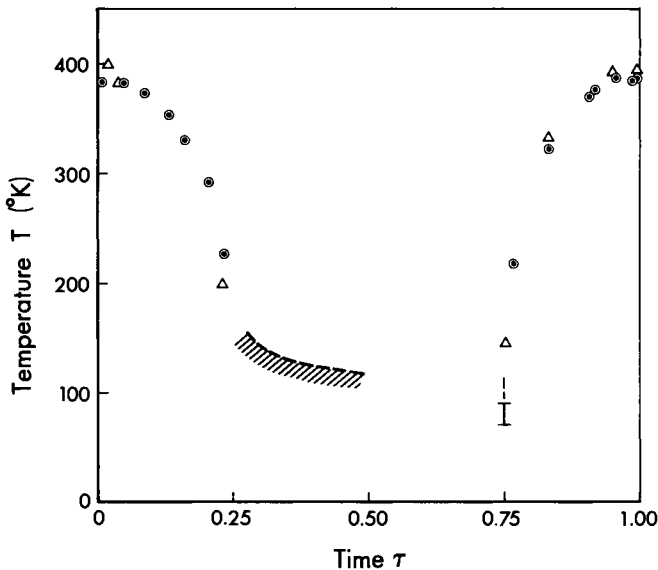
FIG. 1. Experimental data indicating the variation of apparent temperature T with time τ of an area near the lunar equator over a synodic month. Triangles—Sinton (1962); dots in circles—Saari and Shorthill (1967a); I—Low (1965); shaded area—data, Fig. 2.

visible wavelength domains for the purpose of constructing isothermal and isophotic contours (Saari and Shorthill, 1967c). The data shown in Fig. 1 as circled dots were obtained from contour maps at 14 different phase angles and correspond to a region about 20″ arc from the center of the disk (the landing site of Surveyor VI). The exhibited temperatures have been corrected for directional effects which are described in Section V.

In the post-sunset regime, the relative importance of subsurface heat exchange increases rapidly, so that the nighttime thermal response is especially useful in inferring thermophysical soil characteristics. The shaded region in Fig. 1 indicates the range of several sets of post-sunset temperatures exhibited in detail in Fig. 2. The shaded region in Fig. 2 indicates the range of the data obtained from a survey of the midlatitude region of the third-quarter Moon by Wildey *et al.* (1967). Using a mercury-doped germanium photodetector, these workers measured the radiative intensities in equatorial scans in the 8–14 μ region. The resolution of the sensor in these observations was about 26″ arc. If the post-sunset trend is extrapolated to lunar midnight the resulting temperature is approximately 102°K. The figure also shows post-sunset temperatures inferred from measurements by Shorthill and Saari (1965a) (shown as triangles) and from compartments A and B of Surveyor V (Lucas, 1967) (shown by the squares and circles, respectively). The unusually high temperatures
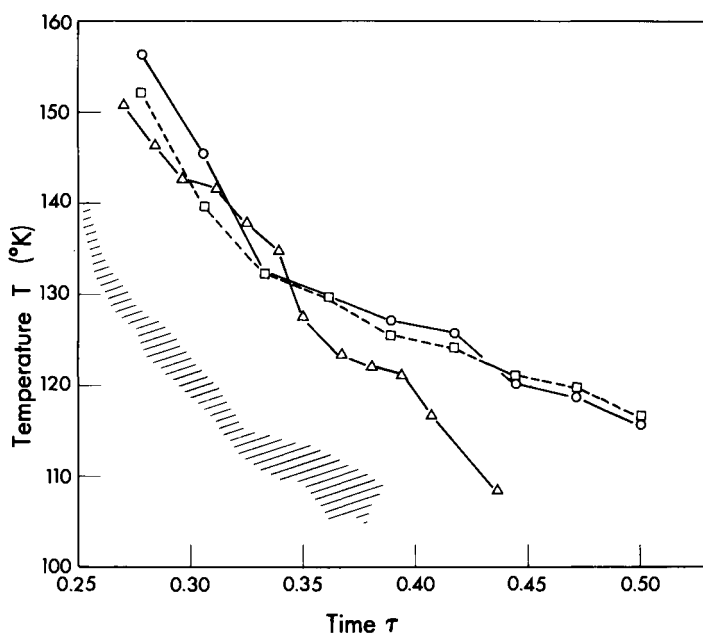
FIG. 2. Experimental data indicating the apparent post-sunset temperature response of an area near the lunar equator. Squares—Surveyor V, compartment A, Lucas (1967); circles—Surveyor V, compartment B, Lucas (1967); triangles—Shorthill and Saari (1965a); shaded areas—Wildey, Murray, and Westphal (1967).

observed by Saari and Shorthill were probably associated with thermally anomalous regions. The Surveyor temperatures were deduced from the thermal history of the outer surface of the compartments using estimated values for the view factors and heat leakage from the interior of the compartments. The data are extremely sensitive to the value assumed for the latter quantity and it should be stressed that the calculated temperatures are preliminary in character. In addition, an uncertainty in the compartment temperatures due to probable telemetry, system, and calibration errors gives rise to an uncertainty of the order of $\pm 9°K$ in the inferred surface temperature. The Surveyor data should probably be regarded as an upper limit since the presence of rocks may produce a substantial enhancement of the apparent brightness temperature of the surface when viewed *in situ* (Roelof, 1968).

Post-midnight temperature measurements require especially sensitive instrumentation. Temperature measurements of the lunar surface just prior to sunrise were made by Low (1965) at the Lunar and Planetary Laboratory of the University of Arizona using a germanium bolometer cooled to 2°K and an interference filter cutting on at 17.5 μ. Low made repeated scans across the

cold limb of the lunar disk and found an average predawn temperature of 90°K, although colder areas were observed with apparent temperatures no greater than 70°K. These data are indicated in Fig. 1 by the vertical line.

Lunation temperature variations alone do not uniquely define the thermophysical parameters of the Moon's surface. It is necessary also to consider variations associated with short-term modulation of the source, such as occur during an eclipse. An especially sensitive measure of the eclipse response is the time variation of the differential energy parameter, P, defined as the difference between the fourth power of the brightness temperature and the solar flux, both quantities being normalized by their preeclipse values (Winter and Saari, 1969). The points used to indicate the time variation of P were taken from the measurements of Saari and Shorthill, obtained during the December 19, 1964 total eclipse. The scanning procedure in this experiment was such that only seven direct measurements of P were available for the disk center. However, evaluation of P at many locations along the lunar equator showed no systematic variation with longitude during the penumbral phase and only a slight variation during totality. Consequently, these values were included as additional data points. The collection of points in Fig. 3 serves to indicate the transient eclipse behavior of P associated with the lunar soil as viewed by a detector overhead.

Infrared measurements during an eclipse have also revealed striking local variations in the cooling rate over the lunar disk which are still the subject of
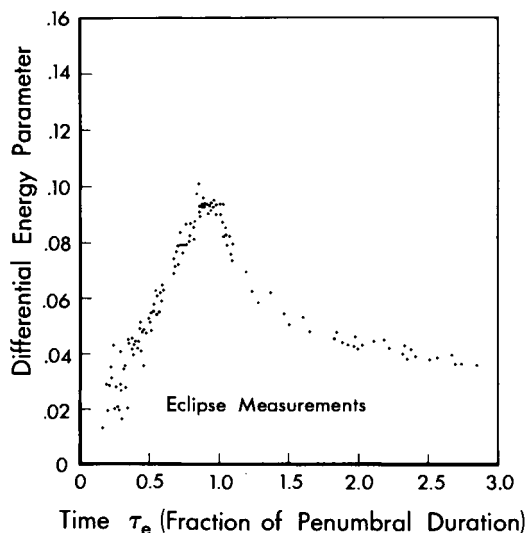
FIG. 3. Measured variation of the differential energy parameter during the December 19, 1964 eclipse. Time τ is normalized to the penumbral duration and is measured from the start of the penumbra (after Winter and Saari, 1969).

considerable speculation. Unfortunately, there are presently available only a few quantitatively reliable cooling histories of selected regions of the surface. The measurements of Pettit (1940), Strong and Sinton (see Sinton, 1962), and Saari and Shorthill (1963) were instructive insofar as they provided a qualitative indication of the response during eclipse cooling. On the other hand, the more recent observations during the eclipse of 18–19 December 1964 by Saari and Shorthill at Boeing Scientific Research Laboratories and Ingrao, Young, and Linsky at Harvard College Observatory are probably more useful for quantitative studies, such as the comparison of various theoretical surface models.

Figure 4 presents anomalous brightness temperatures observed in the vicinity of Tycho and Copernicus during the December 1964 eclipse as reported by
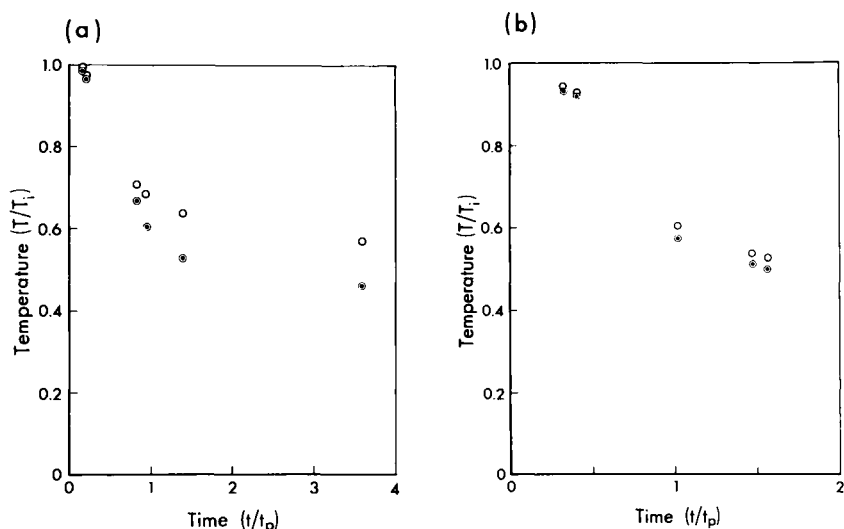
FIG. 4. Observed temporal variations of apparent brightness temperature for the craters Tycho and Copernicus and their environs during the December 19, 1964 eclipse (after Ingrao, Young, and Linsky, 1966). (a) Open circles—crater Tycho; dots in circles—crater Tycho environs. (b) Open circles—crater Copernicus; dots in circles—Copernicus environs.

Ingrao *et al.* (1966). In the figure, apparent temperature T is normalized by the initial temperature T_i and is plotted as a function of time t measured in units of the penumbral duration t_p. The circled dots refer to average temperatures of regions in the environs of the craters. The open circles indicate the cooling response of the interior of the craters. Significant thermal enhancement associated with these features are apparent from the figure.

Observations of the same eclipse by Saari and Shorthill revealed hundreds of local variations in the thermal response of the lunar surface. These observers scanned the lunar disk with a sensor resolution of 10″ arc and obtained detailed maps of radiative intensities throughout the eclipse. These data have been used to construct the transient temperature responses of three selected areas on the lunar disk: the interior of the crater Gambart A, the southeastern portion of Copernicus, and a region to the southwest of Copernicus which appears thermally bland (the "environs") (Winter, 1967). The cooling histories of these areas are shown in Fig. 5, where temperature T is plotted against the fraction of the penumbral time.
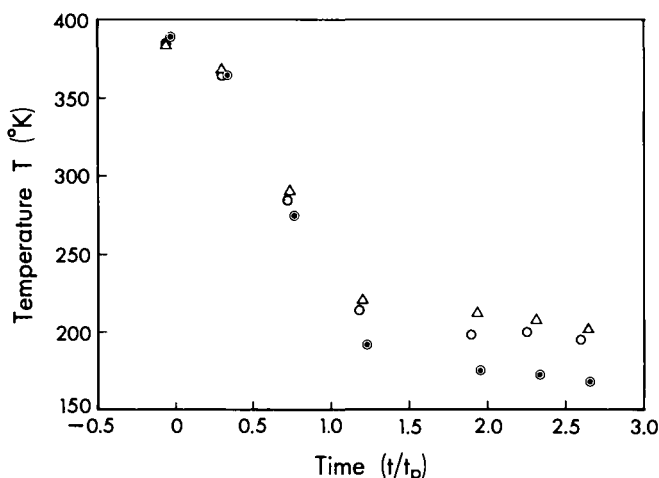
FIG. 5. Observed temporal variation of apparent temperature of the craters Copernicus and Gambart A and a region in their environs during the lunar eclipse of 19 December 1964. Open circles—crater Copernicus, southeast edge; dots in circles—Copernicus environs; triangles— Gambart A.

At this point it is appropriate to inquire whether or not observational data of the sort just presented can be explained by theoretical models based upon a description of the lunar surface as a fine-grained soil. Insofar as the variation of temperature over a lunation is concerned, the daytime temperature is determined almost exclusively by insolation and is not particularly sensitive to the model proposed. On the other hand, the detailed thermal behavior of a region during an eclipse and the lunar nighttime depends critically upon the structure of the soil and its thermophysical properties. Consequently, observed temperatures in the absence of radiation from the Sun enable us to formulate quantitative descriptions of the thermal properties of the upper decimeter. Furthermore, we should expect that any conclusions deduced from lunation data should also

explain the eclipse data. Before describing some recent developments in the calculation of lunar temperature variations, we shall briefly describe the principal mechanisms of heat transfer in heterogeneous material.

III

HEAT TRANSFER IN EVACUATED POWDERS

One may make a distinction between two general types of heterogeneous systems, namely, vesicular or porous media and evacuated powders. An evacuated powder may be defined as a system of solid particles which is surrounded by empty space, while a porous medium may be thought of as a solid material containing an ensemble of discrete spaces. Since the evidence presently at hand indicates that the upper decimeter of the general lunar surface is a fine powder rather than a porous medium, we shall confine the present discussion to powders. We shall also assume that the particles which comprise the lunar soil are opaque at micron wavelengths. It should be noted, however, that if the surface material is composed of particles with dimensions less than a few tens of microns, then the assumption regarding the opacity is probably untenable. The effect of transparency in the infrared will be discussed later.

It is often assumed for a uniform continuum that thermal energy is transferred exclusively by conduction. In other words, the heat flux at an interior point of a homogeneous system is determined by the product of the thermal gradient and a constant thermal conductivity k which is a characteristic of the material. However, in a heterogeneous medium, more than a single process is involved.

Thermal energy transport in an evacuated system of solid opaque particles takes place by solid conduction through the particles, by conduction across areas of contact between adjacent particles, and by radiative exchange between surfaces in view of one another. At the temperatures which are likely to prevail at the surface of the Moon and for particle sizes of the order of a centimeter or less, heat transfer by conduction through the grains will take place just as though the thermal conductivity k of the solid material were infinite. More specifically, an opaque grain of dimension l will be approximately isothermal provided that radiation number $\beta_r = k/\sigma\epsilon T_r^3 l$ is large compared with unity, where T_r is a representative temperature, ϵ is the surface emissivity, and σ is the Stefan-Boltzmann constant.

The thermal response of an evacuated powder is determined principally by contact conduction and radiative exchange acting both in series and parallel. In general, the heat flux in the neighborhood of a given point will depend upon temperature, position, thermal properties of the material, and particle configuration. A comparatively simple analytic description of thermal transport can be achieved by supposing that it is possible to divide the medium into volume

elements, each of which is sufficiently small that a kinetic temperature can be defined within it and, at the same time, is large enough to contain a representative apportionment of the solid and vacuum phases of the medium. Following such an approach, one can define an effective temperature field T, porosity δ, and heat flux $\mathbf{Q}$, which are continuous functions of position and time. This idealization will be referred to as the quasi-continuum model; an alternative particulate model will be discussed in the next section.

In the quasi-continuum model the heat flux $\mathbf{Q}$ at a point within the material can be described by a generalization of Fourier's Law

$$\mathbf{Q} = -k_e(T, \mathbf{x}) \nabla T. \tag{2}$$

This relation defines the effective thermal conductivity as a function k_e of temperature and position which, when substituted into Eq. (2), gives an adequate description of the heat flux in a system of specified boundaries and dimensions. Next, it is assumed that the total flow of heat energy can be treated as a linear superposition of flux due to contact conduction and radiative exchange so that the effective thermal conductivity k_e can be written as the sum

$$k_e = k_c + k_r\,, \tag{3}$$

where the subscripts c and r refer to contributions from conduction and radiation, respectively. Although the form for k_e given by Eq. (3) is convenient, it is not rigorous and it is not difficult to conceive of heterogeneous systems for which linear superposition is incorrect.

The contribution to k_e from thermal contact conduction is difficult to estimate. Although it has long been known that the interface between two solids in contact gives rise to an additional resistance to the flow of heat between them, quantitative estimates of the effect have not been derived until fairly recently. Much of the impetus for a more thorough analysis of contact resistance arose from heat flow problems encountered in nuclear power reactors. Although many of the studies in this field assume the presence of a conducting fluid in interstitial regions, some workers have considered specifically the problem of contact resistance between solids in a vacuum environment (Fenech and Rohsenow, 1963; Clausing and Chao, 1965; Henry and Fenech, 1964; Jeng, 1967; and others).

In the present context, heat flow through contacts between adjacent grains in the lunar soil has been discussed at some length by Watson (1964) and Halajian *et al.* (1967). It is assumed in these studies that the soil structure can be idealized as a cubic array of spheres of equal diameter l. It is then supposed that the contact areas between adjacent spheres in a vertical column are single circular regions of diameter d. The area of each contact is assumed to be determined by the degree of elastic deformation of the spheres due to the weight of

the spheres above the contacts. The constriction ratio d/l is therefore a function of depth below the surface of the powder. For small values of this ratio, the contact resistance across the interface is given by a simple formula. The total contact conductivity k_c in a powder sample of unit cross-sectional area and with a depth of L cm can be calculated in a straightforward fashion. The expression for k_c derived by Watson can be expressed in the form

$$k_c = kLl^{-2/3} \left[\frac{\pi \rho g (1 - \nu^2)}{2E} \right]^{1/3} \left[\sum_{1}^{L/l} j^{-1/3} \right]^{-1}, \tag{4}$$

where k is the thermal conductivity of the material, E is Young's modulus, ν is Poisson's ratio, and g is the acceleration due to gravity. In an elaboration of this same approach, Halajian *et al.* (1967) derive the following expression for k_c which they suggest could be used in Eq. (3):

$$k_c = (1 - \delta) k \left[\frac{\pi \rho g (1 - \nu^2)}{E} \right]^{1/3} x^{1/3}. \tag{5}$$

Here, x is the depth below the surface and δ is the porosity, defined as the volume fraction of the vacuum phase of the system. Both of the above expressions indicate that k_c should depend upon depth but should be relatively insensitive to particle size.

The elastic contact model has produced good agreement with laboratory measurements of thermal conductance across the contact surfaces of metallic specimens subjected to an applied load. On the other hand, it would be somewhat surprising if the same model were to prove equally successful when applied to submillimeter rock particles in a weak gravitational field. Nevertheless, it is difficult to suggest a superior alternative without additional information regarding the nature of contact regions between grains. As a rather extreme alternative, it might be supposed that in the lunar environment the effective contact area between soil grains is more profoundly influenced by the roughness of their surfaces rather than by elastic deformation of the particles under gravity loading. More specifically, suppose that adjacent grains are in contact through n_s microscopic areas which are more or less uniformly distributed in a triangular array over the apparent contact area with a distance of b_s between centers. Further, assume that each of the actual contacts is circular and of identical diameter d_s which is related to the local surface configuration. Under these assumptions it can be shown (Clausing and Chao, 1965) that the contact resistance between adjacent grains is

$$R_s = g(x_s)/k d_s n_s , \tag{6}$$

where the constriction ratio x_s is given by d_s/b_s and where $g(x_s)$ is a constriction alleviation factor which approaches unity for small values of x_s. Certainly,

d_S and n_S will depend in some way upon particle size, material properties, and depth, but the nature of the dependence is not at all obvious. In the case of larger grains, it is conceivable that deformation of a particle as a whole could effectively increase the number of contacts and that deformation of the asperities could increase their effective diameters. The net effect would then be an increase in effective contact area with grain size and depth. On the other hand, this trend could be nullified if the smaller particles of the soil were inherently more rough than larger ones. Since we have so little information to guide us on this point, we shall make the simple working assumption that the product $d_S n_S$ is approximately independent of particle size.

Following Watson, let us next assume that the particle arrangement in the powder can be idealized as a cubic array, so that the total contact resistance of a sample is represented by n^2 resistances in a layer and m resistances in a column. The total thermal resistance of the array is

$$R_C = \frac{1}{n^2} \sum_{j=1}^{m} R_S = \frac{m}{n^2} \frac{g(x_S)}{k d_S n_S} . \tag{7}$$

The contact conductivity in a sample of unit cross-sectional area which is L cm deep can be obtained from Eq. (7) by setting $m = L/l$ and $n = 1/l$:

$$k_C = k d_S n_S / g(x_S) l. \tag{8}$$

The principal difference between this expression and the one displayed in Eq. (4) is that k_C depends now upon the inverse of the particle size.

The microscopic contact model does not give particularly good agreement with measurements of thermal conductance across the interface of metallic solids in the laboratory. The reason for this is that, contrary to assumption, macroscopic contact areas exist and their influence cannot be ignored. The situation may be much the same in the case of rock powder grains. Contacts of both types are probably present and the actual area involved is determined both by gravity loading and by surface configuration. To further complicate the picture, it is possible that differential lateral thermal expansion between the surfaces of grains produces changes in contact areas as the temperature profile within the powder changes.

Evidently, more information regarding the properties of rock powders must be acquired before real progress can be made toward an understanding of the role of contact conduction in the lunar soil. Of considerable value in this connection are certain experimental studies of the thermal properties of powders at low pressures and low temperatures which have recently become available. The work of Wechsler and Glaser (1965), Watson (1964), and Wechsler and Simon (1966) is especially relevant to the problem of heat transfer in the lunar soil. Briefly stated, the results of laboratory measurements seem to indicate that neither

the elastic contact model nor the microscopic contact model is sufficiently realistic to describe the contribution of contact conduction to the flux of heat energy in systems of evacuated powders.

Watson (1964) measured the thermal properties of selected silicate powders enclosed in a vacuum chamber capable of retaining pressures of 10^{-5} to 10^{-6} Torr. The base of the test sample was heated by a temperature bath and the top surface radiated into a cavity maintained at a liquid nitrogen temperature of 77°K. The total thermal conductivity of the material was inferred from the emitted flux as measured with a mercury-doped germanium photodetector. Watson was able to infer values for the contact conductivity of glass microbeads of several different sizes. The experimental results for this material suggest that k_c varies as the inverse of the particle size. Watson also found that values of k_c calculated from theoretical considerations were consistently too large. This finding received further confirmation in the experimental results of Wechsler and Simon (1966). These workers employed the line heat source method, using an ion pumped vacuum system for measurements at pressures of 10^{-8} to 10^{-9} Torr. The thermal conductivities of several silicate powders were inferred from measurements of the rate of temperature rise in the sample. A comparison of observed values of k_c and values calculated from Eq. (7) showed that the theoretical values were too high by a factor of 2 to 5.

Measured values of the contact conductivity of several kinds of powdered material are shown in Fig. 6 as a function of particles size l. The data pertaining to pumice, basalt, quartz powder in the 1 to 10 μ range, and glass microbeads in the 44 to 62 μ range are taken from Wechsler and Simon. The remainder of the measurements are from Watson. The anomalously high value for quartz powder is attributable to the high bulk conductivity of this material at 300°K. An approximate representation of k_c for the several powder materials over the indicated size range is

$$k_c = 3 \times 10^{-8} l^{-1}, \tag{9}$$

shown in the figure as a solid line. This is the same power law suggested by Watson as a fit to the glass microbead data. It should be pointed out, however, that the exponent of l in this equation should not be interpreted as a verification of the independence of contact area and grain size. Both the coefficient and the exponent give representative values of k_c, considering the available data taken together; these choices obviously include effects of material properties as well as particle size. The only firm conclusion which can be drawn at this point is that the theoretical models as presently formulated are too highly idealized to provide anything other than an order of magnitude estimate of contact conduction.

We turn our attention now to the problem of estimating the radiative contribution to the thermal conductivity. There exists a considerable body of literature
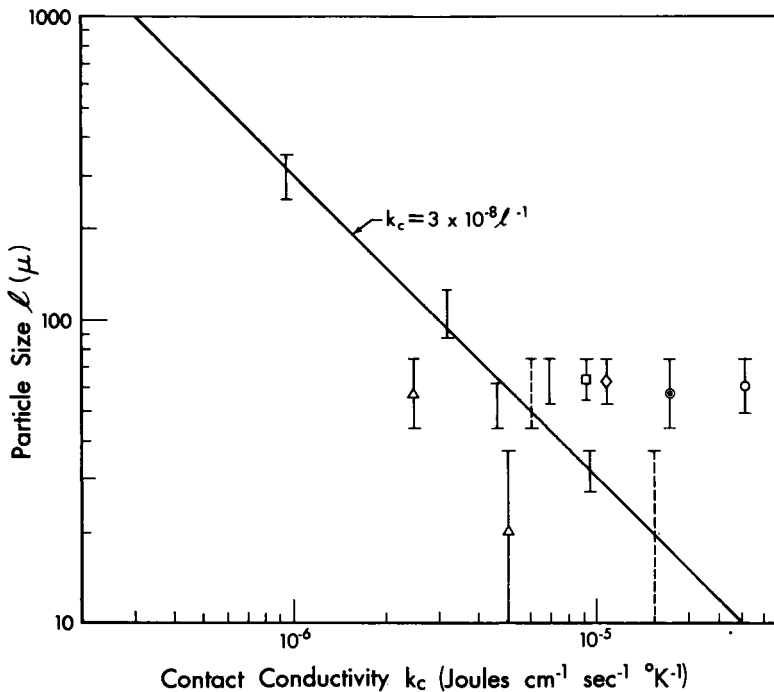
FIG. 6. Measured values of contact conductivity k_c of evacuated powder systems for several ranges of particle size l. Circle—quartz powder; dot in circle—crushed microbeads; dotted lines—basalt powder; triangles—pumice powder; diamond—crushed olivine; square—crushed hornblende; solid lines—microbeads.

on the subject of radiative exchange in powder systems. A comprehensive review of the many developments in this field would not be appropriate here. The major approaches to the problem will be found in discussions by Glaser (1961), Chen and Churchill (1963), and Viskanta (1965); an extensive bibliography is presented by Viskanta and Grosh (1964). In the present context, estimates of the radiative conductivity of the lunar soil have been given by Watson (1964), Clegg *et al.* (1966), Linsky (1966), and Wechsler and Simon (1966).

In spite of extensive investigation, the manner in which system variables influence the transfer of radiant heat energy in an evacuated powder is not well understood. The most general formulation of the problem would certainly include effects of particle placement and configuration, scattering and absorbing properties of the material, and the spectral distribution of the radiation. In order to avoid incapacitating complexity, it has been necessary to resort to highly idealized models to describe the transport of heat by radiation. Most of these

approaches follow more or less the classical treatment of radiative transfer in stellar interiors.

In order to illustrate the basic method, we consider a powder of opaque material in which the only important mechanism of heat transfer is radiative exchange. As pointed out earlier, we need to consider an element of volume sufficiently large to contain a representative apportionment of solid and empty space, yet small enough so that a kinetic temperature can be assigned within it. Most derivations now assume that (1) scattering of radiation within each element is negligible, (2) local thermodynamic equilibrium prevails throughout the medium, and (3) material properties, such as refractive index, emissivity, and photon mean free path, are all independent of wavelength. Considerations are restricted to the steady state and the influence of boundaries is ignored, although Wildey (1967) has recently proposed a method of modifying the traditional approach so that the effect of a boundary might be included.

We assume that radiative transport in a powder is essentially a photon diffusion process which can be described by an appropriate equation for the specific intensity I_ν. In real powders, selective frequency absorption coefficients may preclude the existence of a unique photon mean free path in the infrared. However, it is assumed here that absorption is independent of wavelength and that the grains are opaque so that the mean free path of a photon can be identified with an effective interparticle distance λ_r. Let $\hat{\mathbf{s}}$ be a unit vector in the direction of photon propagation. Using Kirchhoff's law and the assumptions set forth above, the equation of transfer takes the form

$$\lambda_r(\hat{\mathbf{s}} \cdot \nabla) I_\nu = n_r^2 \epsilon B_\nu - I_\nu , \tag{10}$$

where n_r is the refractive index of the material and B_ν is the Planck intensity at frequency ν and temperature T.

When the temperature varies only slightly over a distance equal to the effective interparticle distance λ_r, it is possible to make a Taylor expansion of I_ν about $n_r^2 \epsilon B_\nu$:

$$I_\nu = n_r^2 \epsilon B_\nu - \lambda_r(\hat{\mathbf{s}} \cdot \nabla) n_r^2 \epsilon B_\nu + \lambda_r(\hat{\mathbf{s}} \cdot \nabla)[\lambda_r(\hat{\mathbf{s}} \cdot \nabla)] n_r^2 \epsilon B_\nu - \cdots . \tag{11}$$

It can be seen by inspection that Eq. (11) is a solution of the transfer equation. The specific radiative flux $\mathbf{Q}_\nu$ in direction $\hat{\mathbf{s}}$ due to radiation from all directions is given by

$$\mathbf{Q}_\nu = \int_{\Omega=4\pi} \hat{\mathbf{s}} I_\nu \, d\Omega. \tag{12}$$

Substituting Eq. (11) into Eq. (12), retaining the first three terms of the result, and integrating then over the entire spectrum, we obtain

$$\mathbf{Q} = -(4\sigma\epsilon\lambda_r/3) \nabla n_r^2 T^4. \tag{13}$$

When the refractive index is independent of position and the temperature gradient has only an x-component, Eqs. (2) and (3) can be used to define a radiative conductivity k_r of the form

$$k_r = AT^3, \tag{14}$$

where

$$A = 16n_r^2\epsilon\sigma\lambda_r/3. \tag{15}$$

The factor A is sometimes referred to in the literature as the radiative transfer coefficient.

All that remains is to derive an expression for the effective interparticle distance λ_r. Following developments by Godbee (1966), we begin by simplifying the geometry of the powder system. Consider a sample of the powder contained in a cube measuring L cm on a side whose faces are perpendicular to the axes of a Cartesian triad with the x-axis pointing in the downward direction. Suppose the sample is now sliced into many layers by planes perpendicular to the z-direction. Each layer is sufficiently thin so as to be essentially filled with solid in the z-direction (see Fig. 7a). Now suppose that thermal energy flow
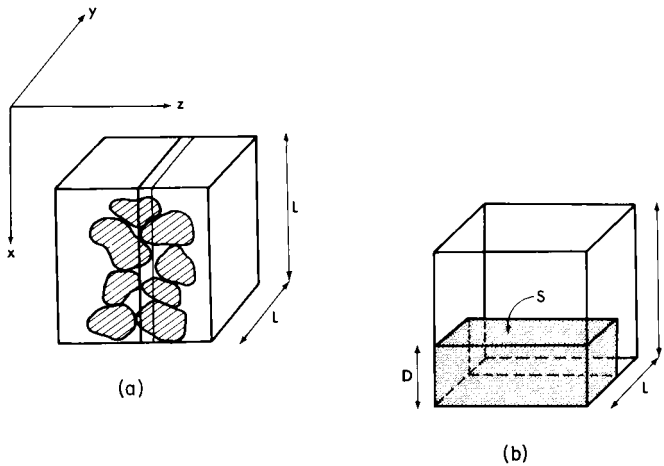
FIG. 7. (a) Geometry of a sample cube of a vacuum-powder system. (b) Simplified body corresponding to the evacuated powder sample shown in (a).

is parallel to the x-direction. If the solid material in each layer is moved to the bottom of the cell along the x-axis, then a consideration of adding resistances in series will show that the effective conductivity of the slice is unaltered. Now suppose the modified system is sliced into very thin layers perpendicular to the x-direction. A consideration of adding conductances in parallel reveals

that the material can now be moved into an essentially vacuum-free rectangular parallelepiped without altering the thermal conductivity. The end result of these maneuvers is the simplified body shown in Fig. 7b. Let D be the extent of the compacted material in the direction of heat flow and let S be the area of its surface parallel to the y–z plane. If l is the representative dimension of a powder grain, then we have, approximately,

$$SD \simeq Sl = (1 - \delta)L^3, \tag{16}$$

where the porosity δ is the volume fraction of the vacuum phase and where D is taken to be approximately equal to the mean particle size l. In a similar fashion, an effective interparticle distance λ_r can be defined by

$$S\lambda_r = \delta L^3. \tag{17}$$

These last two equations give

$$\lambda_r = \delta l/(1 - \delta), \tag{18}$$

and the radiative transfer coefficient becomes

$$A = [16 n_r^2 \epsilon \sigma l \delta / 3(1 - \delta)]. \tag{19}$$

This expression is similar to other forms for A which have appeared in the literature. The form suggested here corresponds most closely to that derived by Godbee (1966).

The experimental efforts referred to earlier have verified, in a general way, the dependence of k_r on the cube of the temperature. In reporting their results, Wechsler and Simon included estimates of the porosity of several powder samples for which the radiative transfer coefficient was inferred. These data are given in Table I, together with measured values of A, for several silicate

TABLE I

Measured and Theoretical Values of A (joules sec^{-1} cm^{-1} °K^{-4})

Material	Size interval (μ)	Porosity (%)	A (meas.)[a]	A (Eq. 19)
Glass beads	44–62	43	2.99×10^{-13}	$(2.53 \pm 0.43) \times 10^{-13}$
Quartz powder	<10	62	3.03×10^{-13}	1.03×10^{-13}
Pumice powder	10–37	69	3.12×10^{-13}	$(3.31 \pm 1.9) \times 10^{-13}$
	44–74	51	3.57×10^{-13}	$(3.88 \pm 0.98) \times 10^{-13}$
Basalt powder	10–37	53	0.88×10^{-13}	$(1.64 \pm 0.94) \times 10^{-13}$
	44–74	51	2.14×10^{-13}	$(3.88 \pm 0.98) \times 10^{-13}$

[a] Wechsler and Simon (1966).

powders. The theoretical values of A given in the table were calculated from Eq. (19). The first number shown is the average for the grain size interval; the differences which follow correspond to the upper and lower limits of the interval. In the calculation of A, the emissivity was assumed to be 0.93 and the index of refraction was assigned the value 1.5. As can be seen from the table, the photon diffusion model gives results which are in fair agreement with experimental data despite the restrictive assumptions upon which the derivation is based.

With the exception of the expression derived by Godbee, the formulas for A given in the literature generally underestimate the radiative thermal conductivity. Most derivations of k_r, including the one shown here, do not allow for partial transparency of the grains. This may account for part of the discrepancy between theory and the experimental results, particularly in the case of silicate powders of very small grain size. If the material is not completely opaque then the photon mean free path may exceed the interparticle distance so that the use of λ_r will lead to an underestimate of A. On the other hand, the actual value of k_r will be reduced by photon scattering processes which have also been excluded from consideration.

A useful measure of the relative importance of radiative exchange and contact conduction is the ratio k_r/k_c. The experimental data for powders in the range 10 to 100 μ show that this ratio may vary from as little as 0.1 to 5 over the temperature range of 150° to 400°K (Wechsler and Simon, 1966). The results of Watson (1964) for microbeads can be used to show that radiative exchange tends to become the dominant heat transfer process when the grain size is of the order of hundreds of microns. The ratio k_r/k_c inferred from Eqs. (9), (14), and (19)
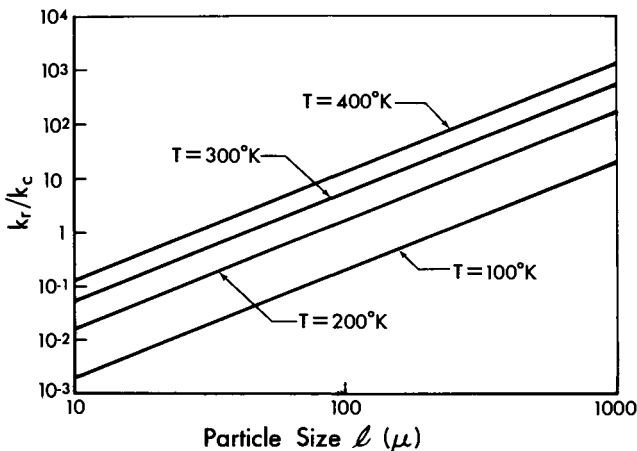
FIG. 8. Ratio of radiative conductivity to contact conductivity as a function of particle size at four different temperatures, as inferred from Eqs. (9), (14), and (19).

is plotted in Fig. 8 as a function of l using the representative values $\epsilon = 0.93$, $n_r = 1.5$, and $\delta = 0.5$. At the lower limit of 100°K, radiative exchange and contact conduction are of comparable importance at l equal to about 280 μ. In the size range 10 to 100 μ the plotted ratios are somewhat less than those quoted by Wechsler and Simon because, in the case of microbeads and pumice powder, Eq. (9) gives an overestimate of k_c.

IV

THEORETICAL ANALYSIS OF SURFACE TEMPERATURE VARIATIONS

Equations describing the thermal response of the lunar surface can now be derived on the basis of the quasi-continuum model using developments of the sort described in the previous section. Since the variations of mass density, effective particle size, and porosity with depth are not well known at present, these parameters will be assumed constant. The rate at which heat energy increases in a unit volume of surface material can be expressed as the product of the mass density ρ, the specific heat c, and the time derivative of the temperature T:

$$\rho c(T) \frac{\partial T}{\partial t} = -\frac{\partial Q}{\partial x}. \qquad (20)$$

At low temperatures the specific heat is a function of T, although there is presently some uncertainty as to the appropriate functional form. Buettner (1963) observed that c should vanish as T approaches zero and found that a plot of averages of specific heats for four igneous rocks was best represented by a dependence of the form $T^{1/3}$. On the other hand, Muncey (1958) and Wechsler and Simon (1966) suggest a linear dependence upon T, while Watson (1964) uses a quadratic form. Figure 9 shows a collection of values for specific heats of igneous rocks (quoted by Buettner), together with data for quartz (Johnson, 1960) and magnesium silicate, calcium feldspar, and silica glass (Goldsmith *et al.*, 1961). For temperatures up to 400°K, a power law of the form

$$c(T) = -0.034T^{1/2} + 0.008T - 0.0002T^{3/2} \qquad (21)$$

is an adequate representation of the data. Although the specific heat is not an especially sensitive function of T in the range 100 to 400°K, definitive calculations of lunar surface temperatures should probably include the variation of specific heat with temperature.

As pointed out in the last section, derivations of the radiative transfer coefficient assume a steady-state condition and apply to regions of material which are far from any boundaries. We now assume that for our purposes these restric-
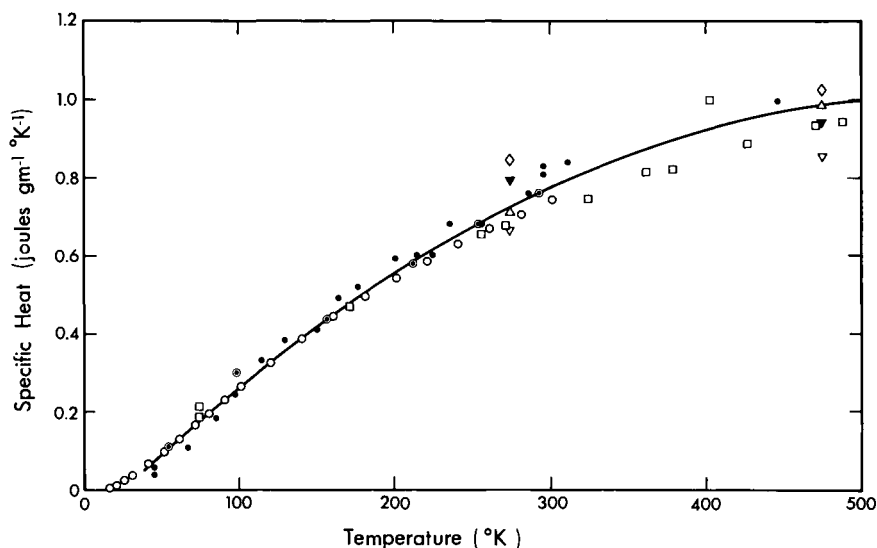
FIG. 9. Variation of specific heat with temperature for several representative rock materials. Heavy dots—magnesium silicate; dot in circles—calcium feldspar; open triangles pointing downward—diabase; pointing upward—diorite; solid triangles—granite; diamond—basalt; squares—silica glass; open circles—quartz.

tions can be suspended and express the heat balance in a unit volume of the lunar soil by the equation

$$\rho c(T) \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} (k_c + AT^3) \frac{\partial T}{\partial x} . \tag{22}$$

The transient temperature profile beneath the lunar surface at selenographic latitude β and longitude λ is found by solving Eq. (22) subject to the boundary conditions

$$\partial T/\partial x \to 0 \qquad \text{as} \quad x \to \infty \tag{23}$$

and

$$(k_c + AT^3) \frac{\partial T}{\partial x} = \epsilon \sigma T^4 - F(t) \qquad \text{at} \quad x = 0, \tag{24}$$

where $F(t)$ is the heat flux into the surface due to insolation and ϵ is the emissivity. For a lunation, the appropriate form for the insolation function is

$$\begin{aligned} F(t) &= (1 - \alpha)(A_0/r_c^2) \cos \beta \cos(\Omega t - \lambda) \qquad \text{when} \quad | \Omega t - \lambda | < \tfrac{1}{2}\pi, \\ &= 0 \qquad \text{when} \quad \tfrac{1}{2}\pi < (\Omega t - \lambda) < \tfrac{3}{2}\pi, \end{aligned} \tag{25}$$

where α is the albedo, A_0 is the solar constant, r_c is the Sun–Moon distance in A.U., and Ω is the synodic angular frequency. When the periodic solution has been found, the temperature variation for an eclipse can be calculated by using the initial condition $T = T(x, 0)$ at (β, λ), together with boundary conditions (23) and (24) in which $F(t)$ is the insolation function appropriate to the eclipse and the selenographic coordinates of the region under consideration. During a good portion of the penumbral phase, the decrease in insolation is roughly linear. On the other hand, if calculations of high accuracy are required, the departure from linearity can be important, as illustrated in Fig. 10. The insolation function shown in the figure corresponds to the central region of the disk for the eclipse of 19 December 1964.
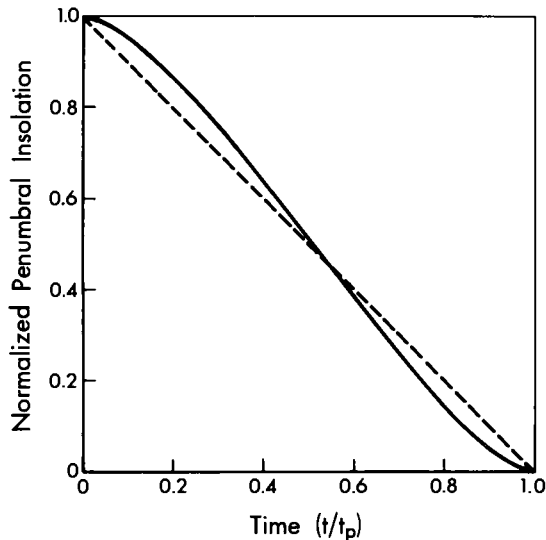
FIG. 10. Variation of the normalized insolation function corresponding to the disk center during the lunar eclipse of 19 December 1964; the dashed line exhibits a linear variation for comparison.

The equations above must be solved by numerical procedures; solutions by finite difference methods have been discussed by Ingrao *et al.* (1966), Linsky (1965), and Halajian *et al.* (1967). Since the values of k_c and A depend upon particle size, once the remainder of the soil parameters are specified it is possible to establish an estimate of l by comparing the theoretically predicted temperature response with that which is observed.

Figure 11 shows theoretically predicted temperature variations of the subterrestrial region of the lunar disk over the course of a synodic month.
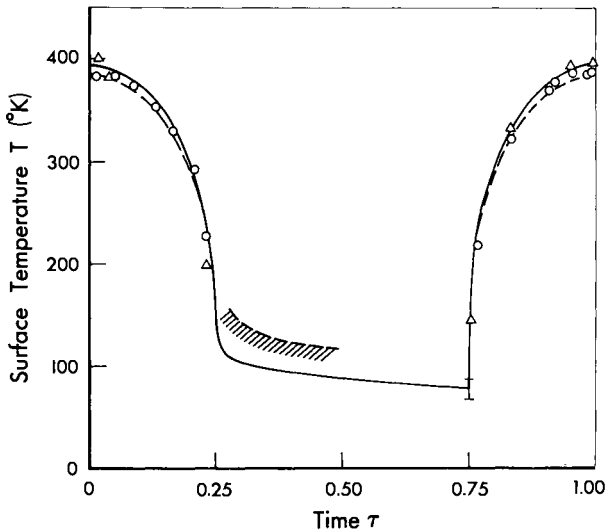
FIG. 11. Theoretically predicted variation of apparent temperature of an area near the lunar equator over a synodic month (quasi-continuum model). Solid line—Wechsler and Simon (1966); dashed line—present formulation.

The solid line in the figure represents the results of a lunation calculation, reported by Wechsler and Simon (1966), based on an analysis by Holland and Ingrao (1966). In this calculation, Wechsler and Simon assigned values to k_C, A, and $c(T)$ in accord with their laboratory results for silicate powders. The dashed line, together with the solid line with which it merges, represents the response predicted with the equations just derived for a particle size of 150 μ. In the calculation, the emissivity ϵ was set equal to unity and the albedo α was chosen to be 0.08 (the value appropriate to the center of the disk). The difference between the two calculations during the lunar daytime is due to different choices for albedo and emissivity. Calculations of a similar nature were performed by Linsky (1966) to investigate the effect of temperature-dependent thermal properties on the radio brightness of the lunar surface. The effective particle size inferred from the microwave study was of the order of 160 to 270 μ.

The pattern of the observed thermal response is reproduced fairly well except in the post-sunset region. Although some of the reported temperatures in this region are probably too high, the discrepancy between theory and experiment is so large as to suggest that additional factors, such as surface relief, must be taken into account.

One alternative to the approach just described is to use a particulate model in which the geometry of the material phase of the medium is idealized in some

way. Since the equations which describe radiative exchange between solid bodies become exceedingly complicated for anything but the simplest shapes, the configuration of the material must be simplified so that the equations describing the physical processes are made tractable. One such model idealizes the soil grains as cubes arranged in a cubic array so that the porosity is 50%. Since the details of this approach have been discussed elsewhere (Winter and Saari, 1969), only the main features will be described here.

As in the quasi-continuum model, considerations were restricted to evacuated powders of opaque material with wavelength-independent thermal properties. Each cube in the array measures l units on a side and is contiguous with its immediate neighbors only along common edges. In order to include the effect of contact conduction, contiguous cubes were imagined to be connected across their common edge by a thermal resistor of value R_S. As a consequence of the opacity assumption, heat energy is transferred only by radiative exchange between cube faces and by interparticle contact conduction. The thermal resistance R_S is related to the contact conductivity by the equation

$$k_c = 2/R_S l. \tag{26}$$

When R_S is independent of depth and particle size, this relation is of the form of Eq. (9) which was established empirically by Watson (1964). In the particulate model, a relationship between R_S and the effective particle size was inferred from a reinterpretation of the laboratory measurements of Watson. In addition, results from Surveyor spacecraft were used to assign representative values to the effective particle size l (40–50 μ) and solid phase mass density (2.7 gm cm^{-3}). The temperature dependence of specific heat was represented by Eq. (21). Surveyor spacecraft experiments suggest that the lunar soil undergoes a transition from a loosely bonded system to a moderately cohesive powder within the upper few millimeters. It is reasonable to expect a corresponding decrease in the thermal resistance with depth. Consequently, the soil model was generalized to allow for a smooth transition in contact resistance starting from surface values appropriate to a noncohesive system (as taken from Watson's measurements) to a value lower by a factor α at depth

$$R_S(x)/R_S(0) = \alpha + (1 - \alpha)\exp(-x^2/x_0^2). \tag{27}$$

The best fit to lunation and eclipse data simultaneously was achieved with $x_0 = 3$ mm and $\alpha = 1/5$, as shown in Figs. 12 and 13. The figures also show the theoretical predictions of the traditional homogeneous continuum surface model for two values of the traditional thermal parameter $\gamma = (k\rho c)^{-1/2}$. These results appear to indicate that no single value of γ will reproduce simultaneously the lunation and eclipse data. Calculations were also performed with a particulate model in which the thermal resistance of the soil remained unchanged with depth.
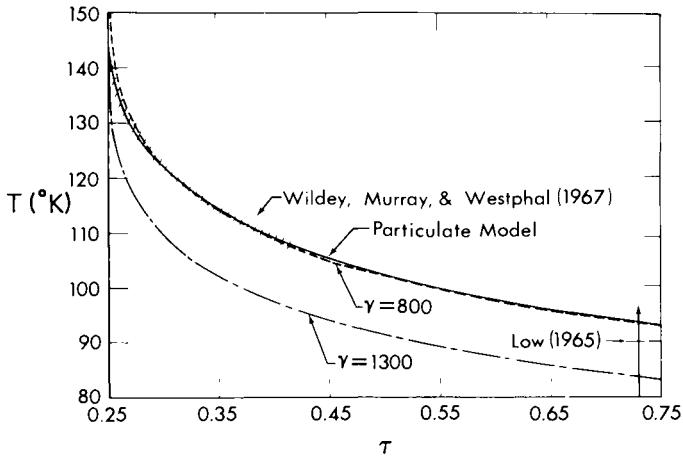
FIG. 12. Predictions of the particulate and constant-properties models compared with observed lunar nighttime brightness temperatures. Time τ is normalized to the synodic period and is measured from local noon (after Winter and Saari, 1969). Dashed area—Wildey, Murray, and Westphal (1967); solid line—particulate model.

The results for uniformly "cohesive" and "noncohesive" soils indicated that lunation and eclipse data cannot be reproduced simultaneously by a model in which the thermal resistance is uniform with depth.

Simultaneous agreement with lunation and eclipse data was achieved primarily because the contact resistance was allowed to vary with depth. Little is known for certain about the actual depth variation of R_S or its relation to other soil parameters. It should be noted that the agreement between the particulate soil model and the data does not constitute a demonstration that R_S is the only depth-dependent parameter. In fact, Jones (1969) has pointed out that a decrease of porosity with depth may have an important effect on the thermal response of the soil. However, the cube model cannot be used to examine this assertion since the porosity is fixed at 50%.

As mentioned earlier, both the particulate and the quasicontinuum models described here are limited by the assumption regarding the opacity of the material. The fact that the grains are not entirely opaque over the relevant spectral domain implies that some heat energy will be transferred by radiation through the grains. In addition to the fact that the mean free photon path will then exceed the grain size, the energy incident upon a distant detector will arise from volume emission, since radiation will no longer be just a surface effect. This aspect of the problem has been discussed at some length by Ulrichs and Campbell (1969) who conclude that volume emission effects may be important in the interpretation of lunar eclipse data and certain laboratory
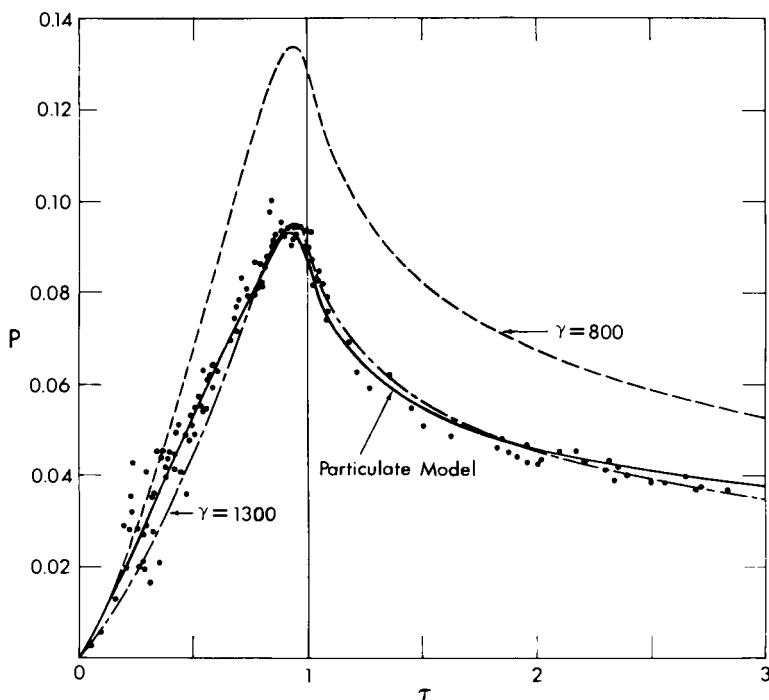
FIG. 13. Predictions of the particulate and constant-properties models compared with measured values of the differential energy parameter from the December 19, 1964 eclipse. Time τ is normalized to the penumbral duration and is measured from the start of the penumbra. The calculations were performed for the disk center (after Winter and Saari, 1969).

powder measurements. The effects of partial transparency and depth-variations of soil properties are the subjects of continuing investigations which will lead ultimately to more complete descriptions of thermal exchange in the lunar soil.

V

THE EFFECT OF SURFACE ROUGHNESS ON APPARENT TEMPERATURE

The upper boundary of the lunar soil has been assumed to be perfectly smooth in the models described thus far. One of the implications of this idealization is that the variation in brightness temperature T over a good portion of the full moon should be described approximately by

$$T = T_s \cos^{1/4} \theta, \tag{28}$$

where T_S is the temperature of the subsolar point and where θ is the angle between the normal to a surface element and the earthward direction. This law follows from Eqs. (24) and (25) if it is assumed that the temperature gradient at the surface can be neglected over the illuminated portion of the disk. However, the early measurements of Pettit and Nicholson (1930) revealed a limb-brightening effect at full moon with the data corresponding more closely to a law of the form

$$T = T_S \cos^{1/6} \theta. \tag{29}$$

The discrepancy between the observed variation and Eq. (28) was ascribed to surface roughness. Pettit and Nicholson offered a qualitative explanation of the effect, using a model consisting of nonconducting spheres uniformly distributed over the lunar surface.

More recent experimental work has verified the existence of the limb-brightening effect. It has been demonstrated, in fact, that the directional emission characteristics of a region on the illuminated lunar surface depend in a rather complicated way upon the elevation angle of the Sun as well as the elevation and azimuth angles of the observer.

Perhaps the most exhaustive study of the directional properties of lunar thermal emission was a by-product of the contouring program, described in Section II, undertaken by workers at Boeing Scientific Research Laboratories. The principal objective of this program was to scan the illuminated disk through a lunation in both the far infrared and visible wavelength domains at a resolution of 8″ arc for the purpose of constructing isothermal and isophotic contours (Saari and Shorthill, 1967b, c). The directional emission properties of the lunar surface were deduced from the data for 19 of 23 phases (Montgomery *et al.*, 1966). One of the findings of this study is summarized in Fig. 14, which shows contours
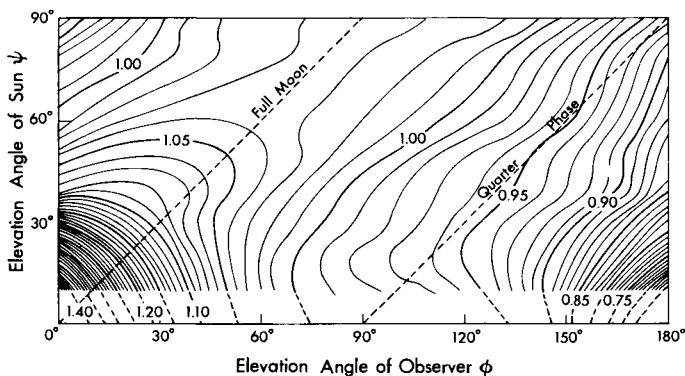
FIG. 14. Contours of the ratio of observed apparent temperature of a point on the thermal meridian to the temperature of a Lambert emitter.

of the ratio of the apparent temperature of a point on the thermal meridian (the great circle through the disk center and the subsolar point) to the temperature of a perfectly smooth diffuse (Lambert) emitter as deduced from Eq. (28). This ratio is a function of elevation angle of the Sun ψ as well as the elevation angle of the observer ϕ. The angle ϕ is measured from 90° to 180° when the lines joining a surface element with the Sun and the observer, respectively, are on opposite sides of the normal to the element (see Fig. 15). The data for
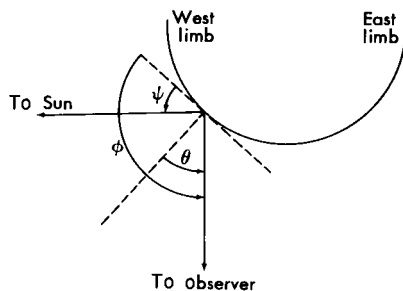
FIG. 15. Definition of elevation angle of sun ψ, elevation angle of observer ϕ, and angle θ corresponding to a point on the thermal meridian.

the subsolar point ($\psi = 90°$) are from Sinton (1962). The temperature ratios plotted in the figure were termed "*D*-factors" (directional factors) by Saari and Shorthill (1967b). The *D*-factors associated with the surface at a given phase correspond to the intersections of the contours with straight lines running obliquely, as illustrated for full moon and quarter phase.

It should be observed that the coordinates (ϕ, ψ) can refer to two different areas on the thermal meridian, one being to the east of the disk center, the other to the west. The exhibited contours correspond to averages of lunar morning and afternoon data. However, the measurements corresponding to *D*-factors within 10° of the left and bottom borders of the diagram in Fig. 14 are subject to large uncertainties. In addition, the contours for ϕ exceeding 130° are of doubtful utility because of the large asymmetry in the morning and afternoon data.

An alternative treatment of the infrared emission data is discussed by Ashby and Burkhard (1967). These workers have developed a relatively simple empirical relation for the apparent emissive power of the illuminated surface as a function of Sun and observer angles. Harrison (1969) has described the application of this relationship to the calculation of the lunar surface thermal flux *in situ*. This work is particularly applicable to the problem of specifying the thermal environment of lunar spacecraft.

Much of the current speculation regarding the directionality of emission

in the infrared has centered upon surface roughness as the major contributing factor. Two general types of surface relief have been considered in this regard, namely, small-scale structure of the order of millimeters or a few centimeters and large-scale irregularities of the order of meters or greater. Low-resolution photographic studies and radar measurements have provided a considerable amount of information regarding large-scale roughness, including discrete relief features, such as craters, and a random surface undulation characterized by an rms slope of 5° to 10°. It is to be noted that the radar signature of the general lunar surface is somewhat less profoundly influenced by roughness on the scale of centimeters or millimeters. On the other hand, small-scale irregularities may play an important role in determining the appearance of the lunar surface at far infrared wavelengths.

One mechanism which may produce and maintain millimeter- or centimeter-scale roughness is continual bombardment of the surface of small meteors and micrometeorites (Bastin, 1966). Microrelief of this order may help to account for the anomalous brightness temperature variation of the subsolar point. An inspection of the contours at the top center of Fig. 14 shows that when the subsolar point is near the center of the lunar disk, the *D*-factor is slightly larger than unity, indicating that the true surface appears to be somewhat warmer than a smooth surface. On the other hand, when the subsolar point is near the limb ($\phi \simeq 0°$ and 180°), the apparent temperature is lower than that of a Lambert emitter.

It has been suggested by Gear and Bastin (1962) that this effect may be explained qualitatively by referring to a corrugated lunar surface model, as shown in Fig. 16. Suppose the surface in the figure is illuminated from directly overhead ($\psi = 90°$). The temperature T_2 at the base of a corrugation will exceed that at the top T_1 because the surface at the bottom receives radiation from material within view at the sides, as well as from the sun. Moreover, if the linear dimensions of the corrugations are of the order of millimeters or a centimeter or so, at the most, then the temperature T_1 can be somewhat less than that of a flat surface of the same material because of radiative losses from the sides near
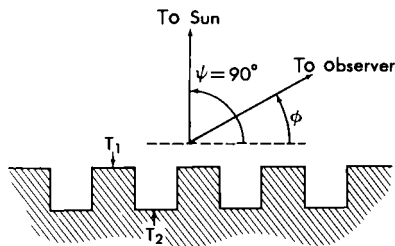
FIG. 16. Corrugated lunar surface model illustrating the effect of small-scale roughness on apparent temperature.

the top of the indentations. Consequently, the emittance of the corrugated surface may exceed that of a smooth surface when viewed from overhead ($\phi = 90°$) but the reverse may be true for observer elevation angles approaching 0° or 180°.

An alternative model has been examined by Buhl *et al.* (1968a) who suggested that full moon limb-brightening and the apparent subsolar point temperature variation may be due to a high density of hemispherical, millimeter-scale craters, such as might be produced by micrometeorites. With regard to the subsolar point variation, for example, these workers argued that significant temperature variations might prevail over the craters even under conditions of normal illumination. The apparent radiation from of a normally illuminated surface with a high density of deep craters differs from that of a smooth surface in that it is somewhat higher in the normal direction and falls below the smooth surface values at small elevation angles, as the observations suggest.

While severe small-scale roughness may account for the anomalous brightness of the subsolar point, it is likely that other forms of roughness become important at low observer angles, as in the case of limb-brightening at full Moon. An interesting explanation of this effect has been offered by Smith (1967) who made an appeal to large-scale irregularities and gross topographic roughness. Smith characterized the large-scale uneveness of the lunar surface as a stochastic variable whose deviation from the mean is a random variable with a Gaussian distribution. A description of the horizontal scale of the irregularities is contained in the autocorrelation function, which is the expected value of the product of surface heights at two locations. Smith assumed that downward heat flow into the illuminated surface could be neglected and that each element of the random surface was a Lambert emitter at infrared wavelengths. In order to simplify the analysis, it was further assumed that radiative exchange between surface area elements could be ignored. This limits the validity of the treatment to large-scale relief characterized by comparatively gentle slopes. According to this model, a brightness enhancement near the limb at full moon is due to the existence of a set of surface elements which are inclined for more efficient radiation in the Sun–Earth direction. Smith derived an expression for the radiant emittance of the surface viewed at angle ϕ which depends upon the rms slope. The theoretical results for rms slopes of 10° to 20° are shown by the solid lines in Fig. 17, in which the relative radiant emittance ($T^4/T_S{}^4$) at full moon is plotted as a function of elevation angle of the Sun, ψ. Shown in the same figure are experimental results reported by Ingrao *et al.* (1966) and Six *et al.* (1967). The data assembled by the latter are well represented by a law of the form

$$T^4 = (324.2 + 72.6 \sin \psi)^4. \tag{30}$$

The best fit to the experimental results is achieved by the theoretical curve corresponding to an rms slope of 20°. Although this value is somewhat higher
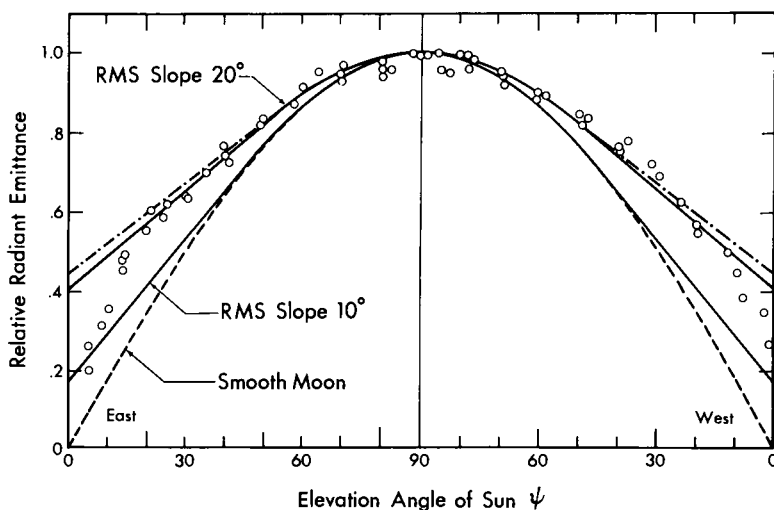
FIG. 17. Relative radiant emittance at full moon as a function of elevation angle of the Sun, ψ. Circles—Ingrao, Young, and Linsky (1966); dot-dash line—Six, Montgomery, Saari, and Shorthill (1967).

than suggested by an analysis of radar returns, it appears to represent the full-moon data reasonably well. On the other hand, an attempt to account for the variation of the subsolar point brightness with phase by the same model results in a value for the rms slope which is so large (60°–70°) that the assumptions upon which the analysis is based are probably no longer tenable.

Another factor which may contribute to the directional effect is the presence of rocks on the surface. Photographic evidence from the Surveyors, Luna IX, and the Lunar Orbiter spacecraft indicates that the general lunar surface may be strewn with debris which has the appearance of rock. Although some of these structures may be clumps or agglomerations of powder (Hapke, 1967), if an appreciable fraction are actually solid, their thermal behavior during a lunation and an eclipse will differ significantly from that of the lunar soil. The photographic evidence indicates a distribution of rocks over a wide range of sizes, from a few millimeters to several meters in the vicinity of craters.

The effect of rocks on both the optical properties and the apparent thermal response of the lunar surface was first examined by Hopfield (1967). Quantitative assessments have recently been reported by Roelof (1968) who examined the thermal behavior of both centimeter- and meter-size lunar rocks which were idealized as cubes of basaltic material resting on the surface. It was assumed that their exposed surfaces received radiation from the lunar surface and from the sun and that the cubes were in poor thermal contact with the soil. In view of this last assumption, the results should be interpreted as providing approxi-

mate upper bounds for the influence of rocks on apparent surface brightness temperatures. During most of a lunation, centimeter-size rocks remain approximately isothermal and are slightly cooler than the general surface. An exceptional situation arises at sunset and sunrise, when small rocks may exhibit considerably higher temperatures than the lunar soil. Figure 18 shows plots of differences in
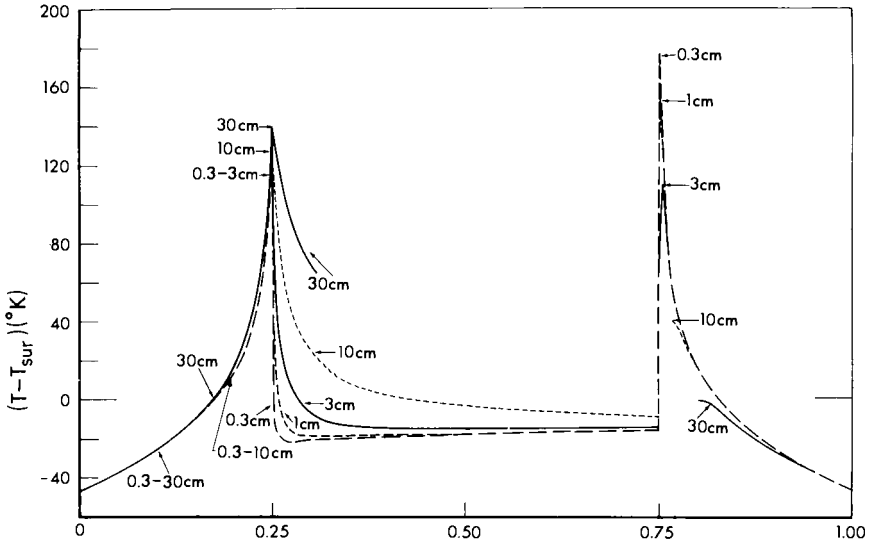
FIG. 18. Difference between small rock temperature T and surface temperature T_{sur} of a region on the lunar equator during a synodic month (after Roelof, 1968).

temperature between rocks of various sizes and the surface for a region on the equator during a lunation. The differentials will be reduced somewhat by conductive heat exchange with the soil and partial imbedding. In the case of meter-size rocks, the isothermal approximation no longer holds and each face of the cube possesses a unique thermal history during a lunation, as illustrated in Fig. 19. Perhaps the most important effect is the exceptional brightness of the west face at sunset and the east face at sunrise. Roelof reached the general conclusion that the net effect on the brightness temperature inferred by a remote observer will usually be small. On the other hand, thermal emission from surface debris may assume some importance when the detector-response function is such as to give more weight to emission from the debris than from the soil or when the elevation angle of the observer is close to 0° (or 180°). Under the latter circumstance, i.e., as the angle between the normal to the surface and a ray to the detector increases, a correspondingly greater percentage of the soil will be occulted by debris. Such a situation would prevail at full moon near the limbs where the rock-surface temperature differentials are especially
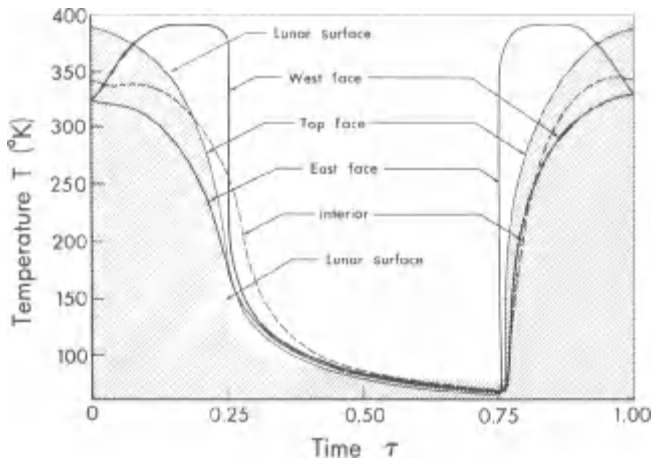
FIG. 19. Thermal history of a 1 meter basaltic cube on the lunar equator during a synodic month (after Roelof, 1968).

pronounced. These considerations suggest that debris in the form of solid rocks on the lunar surface may contribute to the limb-brightening effect at full moon.

VI

THERMAL ANOMALIES ON THE LUNAR SURFACE

During the lunar eclipse of 13 March 1960, Shorthill *et al.* (1960) observed that the ray crater Tycho cooled more slowly than its immediate environs. This unexpected result was verified by Sinton (1960) during the 5 September 1960 eclipse. These discoveries prompted the speculation that high-resolution scans during an eclipse and a lunation might provide more evidence of anomalous thermal behavior which, in turn, could provide further clues to the nature of the lunar surface structure. Subsequent eclipse observations by Saari and Shorthill (1963) and Ingrao *et al.* (1966) revealed that several ray craters seemed to show a significantly slower thermal response than the general lunar surface. Some of these same features also were found to be anomalously warm just after sunset (Murray and Wildey, 1964).

Observations of the 19 December 1964 eclipse by Saari and Shorthill (1965) revealed pronounced nonuniformities in cooling rate over nearly all of the lunar disk. Qualitatively, the thermal heterogeneity is perhaps best illustrated by the infrared image of the eclipsed Moon which is shown in Fig. 20. It can be seen from the image that the major ray craters, such as Tycho, Copernicus, and
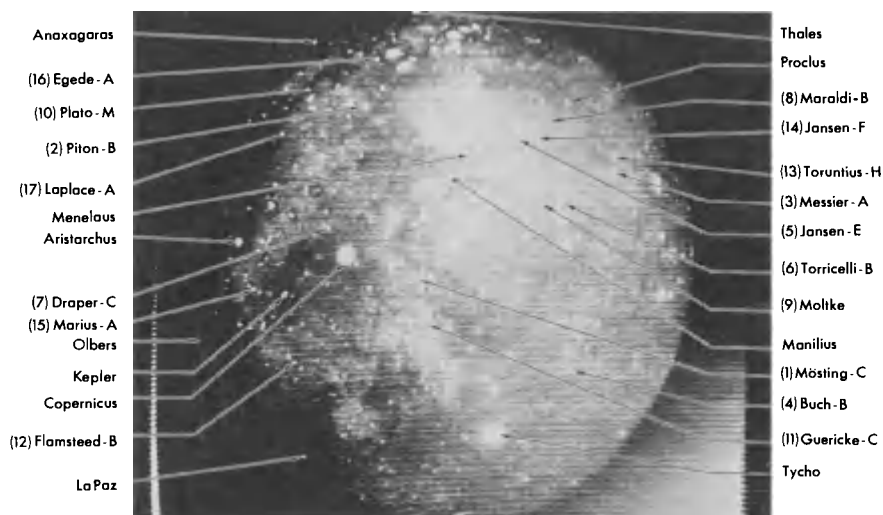
FIG. 20. Infrared images of the Moon as reconstructed from the data during the umbral phase of the eclipse of 19 December 1964 (after Saari and Shorthill, 1967b).

Aristarchus are unusually bright. In addition, there appear to be literally hundreds of anomalous regions or "hot spots," some of which are not associated with ray craters. It is also apparent from the image that the temperature of some of the maria or portions thereof is slightly elevated in comparison with adjacent regions.

Several hundred thermal anomalies have been catalogued and identified for the purpose of studying the relationship between hot spots and visual features (Shorthill and Saari, 1965b). Table II presents a preliminary classification based upon 330 anomalous regions, over 90% of which are visually bright in some respect at full moon. On the other hand, over 10% of the anomalies do not

TABLE II[a]

PRELIMINARY CLASSIFICATION OF THERMAL ANOMALIES ON THE ECLIPSED MOON

Ray craters	19.4%
Craters with bright interiors at full moon	41.8%
Craters with bright rims at full moon	23.3%
Bright areas with much smaller craters	3.6%
Bright areas associated with features like ridges	3.9%
Bright areas not associated with any features	1.2%
Craters not bright at full moon	0.6%
Position unidentified or questionable	6.3%

[a] After Saari and Shorthill (1967b).

appear to be directly associated with crater features; some appear in bright areas which may contain a small crater, such as Deslandres, while others are associated with rilles, such as Rima Hadley and Hyginus. Another surprising feature of hot spots is the nonuniformity of their distribution over the disk. Nearly two-thirds appear in maria while the remaining one-third are found in highland regions. The greatest concentrations of thermal anomalies are to be found in Mare Tranquillitatis (8 per 10^5 km^2) and in a region between Aristarchus and Kepler (6 per 10^5 km^2).

In order to make proper assessments of possible mechanisms which produce thermal anomalies it is necessary to use quantitative descriptions of the thermal response, such as the march of apparent temperature or radiant emittance with time. Unfortunately, not many "cooling curves" are available which are sufficiently accurate for such a purpose. However, the data presented in Figs. 4 and 5, drawn from the observations of Ingrao *et al.* (1966) and from Saari and Shorthill (see Winter, 1967), give a fair indication of the thermal history of a hot spot, and any plausible model for a mechanism ought to be able to reproduce the general features of these data. In addition to "cooling curves," other types of data presentation have proved useful, including detailed thermal contouring of hot spot regions (Saari and Shorthill, 1967c) and the construction of infrared images (Salisbury and Hunt, 1967).

In spite of the wealth of observational data, the factors which give rise to thermal anomalies have not yet been definitely identified. A number of plausible mechanisms of different character have been suggested, but no single one has yet emerged as preferable to all others. It is more likely that several factors are involved in producing anomalies, with some operating more effectively than others in certain regions. Amongst the explanations which have been considered in the literature are the following: (1) emissivity variations, (2) differences in infrared transparency, (3) subsurface (internal) heat sources, (4) exposure of bare rock, (5) larger soil particle size, (6) centimeter-scale surface roughness, and (7) surface rocks. The first of these possibilities is one of the least likely, as has been shown by Saari and Shorthill (1963), Fudali (1966), and Salisbury and Hunt (1967). A region of the surface covered by material whose emissivity is intrinsically lower than its environs will cool less rapidly. However, in order to produce temperature differentials of the order of tens of degrees Kelvin, the emissivities of the contrasting areas must differ by a factor of two. Differences of this order are not consistent with emissivity variations observed in terrestrial rock.

Differences in infrared transparency of lunar materials may also produce variations in the apparent thermal response of the surface (Buettner, 1963). While such a mechanism cannot be ruled out altogether, it is unlikely to play a major role in the more prominent anomalies because excessively large differences in transparency would be required.

It has been stated in the literature that the third alternative, namely, subsurface heat sources, is an unlikely explanation since anomalous areas are not unusually warm during the lunar day. However, it may be premature to exclude this possibility in all cases. Rough calculations have shown that an internal heat source which produces an enhancement of only a few degrees during the lunar daytime can give rise to a hot spot with a temperature enhancement of several tens of degrees Kelvin during a lunar eclipse. It also appeared from the calculations that such an area would tend to remain warmer than its environs, by ten to twenty degrees, during the lunar nighttime.

One of the more obvious explanations of a hot spot is an expanse of bare rock or a substratum of compact rock with an overlaying thin layer of dust (Saari and Shorthill, 1963). It is conceivable that a morphology of this sort could contribute to the anomalous response of certain portions of young, ray craters. On the other hand, it is highly unlikely that anomalous features consist entirely of bare rock, since the temperature enhancement would then be considerably greater than is observed. A possible exception to this last statement is the crater Mösting C. The data associated with this feature suggest a temperature enhancement of 157°K when special assumptions are made regarding the appropriate areal correction (Shorthill and Saari, 1965b). After a crater is formed, meteoritic infall will tend to cover any exposed rock substrata with blanketing and pulverizing debris. On the other hand, downslope migration of meteoritic debris may serve to retard the accumulation of dust in regions characterized by high-angle slopes (Fudali, 1966). One might expect such a mechanism to operate most effectively on steep interior rims of youthful craters. Some of the high-resolution photographs taken by the Lunar Orbiters suggest slumping of material from the interior rims of certain thermally anomalous craters, such as Gambart C.

An anomalous eclipse response could also indicate a coarser soil with a larger effective grain size. For example, a twofold increase in the effective particle size can produce a 5 to 10% increase in the surface temperature during totality. A slight shift in the effective particle size may contribute to the general diffuse enhancement of certain maria in contrast with upland regions. If this is true, the implication would be that upland areas are covered with finer material than younger maria which have not been exposed so long to blanketing and pulverizing by meteoritic dust. It is interesting to note in this connection that recent laboratory studies suggest that the effective particle size of a powder system may be a key factor in determining its albedo (Adams and Filice, 1967). It is not unlikely, therefore, that a shift in mean particle size may also help to account for albedo differences between uplands and maria (Adams, 1967).

Another plausible mechanism for changing the apparent thermal response of the lunar surface is centimeter-scale surface roughness (Bastin, 1965; Winter, 1966a, b; Buhl *et al.*, 1968b). The effect of small-scale roughness on

the apparent temperature of the surface has already been discussed in connection with directional effects on the illuminated Moon. The observation to be made in the present context is that a surface which is rough on a centimeter scale can appear to be brighter than a plane surface of the same material even when a steady state does not prevail. Severe disorganization of the surface in the vicinity of a feature whose origin is catastrophic may thereby contribute to an apparently anomalous temperature decline during eclipse cooling. The roughness hypothesis was originally put forward to account for the anomalous response associated with the floors and immediate exteriors of craters. Quantitative estimates of the influence of roughness on temperature enhancements have demonstrated that the magnitude of the effect is sufficient for small-scale roughness to qualify as a contributor in some regions.

It is also possible that thermal anomalies are caused, in part, by a profusion of rocks and surface debris. High-resolution Lunar Orbiter photographs have revealed large quantities of what appear to be meter-size rubble scattered over the floors of craters and the immediate environs of their rims. It seems likely that if the size distribution of surface debris in a crater area is rich in meter-scale rocks, then the apparent cooling response of the area could be profoundly affected. At this writing, a quantitative estimate of the effect of large rocks during a lunar eclipse is unavailable. However, certain trends in the calculations of Roelof (1968) (discussed earlier in connection with lunation cooling) suggest that large-scale debris might contribute significantly to apparent thermal enhancements during an eclipse. Roelof has calculated the thermal response of small solid rocks on top of the lunar soil during a lunar eclipse with a penumbral duration of one hour. The size of the rock considered in the study of eclipse

TABLE III[a]

PERCENTAGE THERMAL ENERGY ENHANCEMENTS DUE TO ROCKS ON SURFACE[b]

Rock size range (cm)	$t = 0$, $T_S = 393°K$		$t = 1$ hr, $T_S = 222°K$		$t = 3$ hr, $T_S = 160°K$	
	I*	III*	I	III	I	III
0.3–1	−0.38	−1.46	0.42	1.49	−0.19	−0.89
1–3	−0.45	−0.69	1.62	2.28	1.05	1.42
3–10	−0.64	−0.40	4.53	2.71	4.56	1.04
(10–30)	−0.77	−0.20	7.63	1.86	19.06	5.30
0.3–10	−1.48	−2.55	6.57	6.48	5.42	1.57
(0.3–30)	−2.24	−2.75	14.20	8.35	24.48	6.89

[a] After Roelof (1968).
[b] Using rock-size distributions from Surveyor I (unrevised) and III.

cooling was limited to 30 cm because the numerical techniques employed were not adequate to describe the transient temperature field in larger objects during drastic changes in insolation. The effect of a distribution of rock sizes was investigated by calculating the transient overhead brightness temperature of a surface strewn with rocks with two different size distributions. One distribution was inferred from Surveyor III photographs, while the other was the preliminary (unrevised) distribution for Surveyor I. The latter is less steep and richer in large-scale debris. A comparison of the results of the two distributions indicates a trend which lends credibility to the assertion that meter-scale debris can produce a hot spot. Table III shows the percentage thermal energy enhancement of rubble-strewn surfaces for the two size distributions. Results are displayed for three periods of time during an eclipse with a penumbral time of one hour. Although neither of the two distributions considered can account for the larger observed differentials, the distribution which is richer in larger debris produces a significantly higher percentage enhancement even with a particle size cutoff at 30 cm. It is plausible, then, that a distribution which includes meter-size debris could reproduce the observed enhancements of rubble-strewn regions of craters.

Relevant to these speculations are the observations of Allen and Ney (1969) who measured the lunar nighttime cooling of the craters Tycho, Copernicus, and Aristarchus at wavelengths between 3 and 14 μ. All three features, which are prominent hot spots during an eclipse, exhibited persistent nighttime color temperatures of the order of 200°K. The data were interpreted by Allen and Ney as indicating that as much as 10% of the surface might be covered by meter-sized debris in good thermal contact with substrata of the regolith. This work further substantiates the hypothesis that eclipse anomalies associated with youthful craters are primarily due to surficial boulder fields.

It is an interesting fact that delay-frequency analysis of radar backscattering from the Moon has revealed that some features which are thermally enhanced during an eclipse, such as Tycho and Heinsius A, also produce anomalous radar returns at centimeter wavelengths. The relatively stronger radar returns from these features have been variously attributed to localized areas of exposed compacted rock with a high dielectric constant and to severe surface roughness on a scale of meters (Thompson and Dyce, 1966; Pettengill and Thompson, 1968). Since such surfaces can be expected to exhibit anomalous responses at infrared wavelengths, one might hope to find some correspondence between the details of contour maps of radar scattering and infrared emission. However, no simple correspondence between the contours is evident (Shorthill and Saari, 1965b). Moreover, some craters which are anomalous radar scatterers, such as Eratosthenes and Posidonius, are almost completely thermally bland during an eclipse. In addition, the upland regions of the Moon reflect $1\frac{1}{2}$ to 2 times as much power to a radar receiver as the maria, whereas at infrared wavelengths,

it is the maria which exhibit enhancements. The reasons for these differences are not completely understood.

The objective of the present survey has been to outline the general status of the subject at the outset of the Apollo era. Despite the very considerable progress during the 1960's, a number of questions regarding the structure and the thermal properties of the lunar surface are yet unanswered. However, with manned excursions to the Moon, our knowledge of the lunar surface and its thermophysical properties should expand at a rapid pace, and many of the uncertainties which pervade the subject at present will be resolved.

Note Added in Proof

After this chapter was completed, further theoretical and observational work at infrared wavelengths established that centimeter- and meter-scale roughness can control the apparent thermal response of the lunar surface under many circumstances. In fact, recognition of the importance of surface relief has led to additional color temperature observations and more detailed theoretical descriptions of the temperature variations associated with relief features. Various aspects of these topics have been treated by Adorjan (1971), Allen (1971a, b), Buhl (1971), Bastin and Gough (1969), Winter (1970), Winter and Krupp (1971), and others. In summary, it is presently believed that intermediate-scale craters, both sharp and subdued in form, are responsible for most of the infrared directional characteristics of the sunlit lunar surface. During the lunar nighttime or an eclipse, on the other hand, positive relief in the form of surficial rocks and boulders partially determines the apparent temperature response at a given wavelength, particularly near the limb of the Moon. Finally, it is now generally agreed that the majority of thermal anomalies which appear during the lunar nighttime and an eclipse is produced by exceptionally dense populations of surface rocks. Details of the thermal characteristics of lunar surface roughness will be found in a survey by Winter *et al.* (1972).

References

Adams, J. B. (1967). *J. Geophys. Res.* **72**, 5717.
Adams, J. B., and Filice, A. L. (1967). *J. Geophys. Res.* **72**, 5705.
Adorjan, A. S. (1971). AIAA 6th Thermophys. Conf., Pap. No. 71-480.
Allen, D. A. (1971a). *The Moon* **2**, 320.
Allen, D. A. (1971b). *The Moon* **2**, 435.
Allen, D. A., and Ney, E. P. (1969). *Science* **164**, 419.
Ashby, N., and Burkhard, D. G. (1967). *NASA Contract. Rep.* NASA CR-61481.
Bastin, J. A. (1965). *Nature (London)* **207**, 1381.
Bastin, J. A. (1966). *Nature (London)* **212**, 171.

Bastin, J. A., and Gough, D. O. (1969). *Icarus* **11**, 289.
Buettner, K. J. K. (1963). *Planet. Space Sci.* **11**, 135.
Buhl, D. (1971). *J. Geophys. Res.* **76**, 3384.
Buhl, D., Welch, W. J., and Rea, D. G. (1968a). *J. Geophys. Res.* **73**, 5281.
Buhl, D., Welch, W. J., and Rea, D. G. (1968b). *J. Geophys. Res.* **73**, 7593.
Chen, J. C., and Churchill, S. W. (1963). *Amer. Inst. Chem. Eng. J.* **9**, 35.
Clausing, A. M., and Chao, B. T. (1965). *J. Heat Transfer* **87**, 243.
Clegg, P. E., Bastin, J. A., and Gear, A. E. (1966). *Mon. Notic. Roy. Astron. Soc.* **133**, 63.
Epstein, P. S. (1929). *Phys. Rev.* **33**, 269.
Fenech, H., and Rohsenow, W. M. (1963). *J. Heat Transfer* **85**, 15.
Fudali, R. F. (1966). *Icarus* **5**, 536.
Gear, A. E., and Bastin, J. A. (1962). *Nature (London)* **196**, 1305.
Glaser, P. E. (1961). *In* "International Developments in Heat Transfer," Part IV, pp. 829–837. Amer. Soc. Mech. Eng., New York.
Godbee, H. W. (1966). Ph.D. Thesis, Georgia Institute of Technology; also Oak Ridge Nat. Lab. Rep. ORNL-3510 (1966).
Goldsmith, A., Waterman, T. E., and Hirschhorn, H. J. (1961). "Handbook of Thermophysical Properties of Solid Materials," Vol. III. Macmillan, New York.
Halajian, J. D., Reichman, J., and Karafiath, L. L. (1967). "Correlation of Mechanical and Thermal Properties of Extraterrestrial Materials," Grumman Res. Dep. Rep. RE-280.
Hapke, B. (1967). *Icarus* **6**, 254.
Harrison, J. K. (1969). *Int. J. Heat Mass Transfer* **12**, 689.
Henry, J. J., and Fenech, H. (1964). *J. Heat Transfer* **86**, 543.
Holland, J. T., and Ingrao, H. C. (1966). "A Study of the Thermal Response of the Lunar Surface at the Landing Site During the Descent of the Lunar Excursion Module (LEM)." NsG 64-60, Tech. Rep. Harvard College Observatory, Cambridge, Massachusetts.
Hopfield, J. J. (1967). *In* "Physics of the Moon," Sci. Technol. Ser. No. 13, pp. 101–120. Amer. Astronaut. Soc., Tarzana, California.
Ingrao, H. C., Young, A. T., and Linsky, J. L. (1966). *In* "The Nature of the Lunar Surface" (W. N. Hess, D. H. Menzel, and J. A. O'Keefe, eds.), pp. 185–211. Johns Hopkins Press, Baltimore, Maryland.
Jaeger, J. C. (1953). *Austr. J. Phys.* **6**, 10.
Jaeger, J. C., and Harper, A. F. A. (1950). *Nature (London)* **166**, 1026.
Jeng, D. R. (1967). *J. Heat Transfer* **89**, 275.
Johnson, V. J. (1960). "A Compendium of the Properties of Materials at Low Temperatures; Part II. Properties of Solids," WADD Tech. Rep. 60-56.
Jones, B. P. (1969). *Progr. Astronaut. Aeronaut.* **21**, 469–487.
Krotikov, V. D., and Shchuko, O. B. (1963). *Sov. Astron.—AJ* **7**, 228.
Krotikov, V. D., and Troitsky, V. S. (1963). *Sov. Phys.—Usp.* **6**, 841.
Linsky, J. L. (1965). "A Computer Program to Solve the Heat-Conduction Equation in the Lunar Surface for Temperature-Dependent Thermal Properties," NsG 64-60, Sci. Rep. No. 7. Harvard College Observatory, Cambridge, Massachusetts.
Linsky, J. L. (1966). *Icarus* **5**, 606.
Low, F. J. (1965). *Astrophys. J.* **142**, 806.
Lucas, J. W. (1967). Personal communication.
Montgomery, C. G., Saari, J. M., Shorthill, R. W., and Six, N. F., Jr. (1966). "Directional Characteristics of Lunar Thermal Emission," Boeing Doc. D1-82-0568; Brown Eng. Co. Tech. Note R-213.

Muncey, R. W. (1958). *Nature (London)* **181**, 1458.
Muncey, R. W. (1963). *Aust. J. Phys.* **16**, 24.
Murray, B. C., and Wildey, R. L. (1964). *Astrophys. J.* **139**, 734.
Pettengill, G. H., and Thompson, T. W. (1968). *Icarus* **8**, 457.
Pettit, E. (1940). *Astrophys. J.* **91**, 408.
Pettit, E., and Nicholson, S. B. (1930). *Astrophys. J.* **71**, 102.
Roelof, E. C. (1968). *Icarus* **8**, 138.
Saari, J. M., and Shorthill, R. W. (1963). *Icarus* **2**, 115.
Saari, J. M., and Shorthill, R. W. (1965). *Nature (London)* **205**, 964.
Saari, J. M., and Shorthill, R. W. (1967a). Personal communication.
Saari, J. M., and Shorthill, R. W. (1967b). *In* "Physics of the Moon," Sci. Technol. Ser. No. 13, pp. 57–99. Amer. Astronaut. Soc., Tarzana, California.
Saari, J. M., and Shorthill, R. W. (1967c). *NASA Contract. Rep.* **NASA CR-855**.
Salisbury, J. W., and Hunt, G. R. (1967). *Icarus* **7**, 47.
Shorthill, R. W., and Saari, J. M. (1965a). *Ann. N. Y. Acad. Sci.* **123**, 772.
Shorthill, R. W., and Saari, J. M. (1965b). *Science* **150**, 210.
Shorthill, R. W., Borough, H. C., and Conley, J. M. (1960). *Publ. Astron. Soc. Pac.* **72**, 481.
Sinton, W. M. (1960). *Lowell Obs. Bull.* **5**, 25.
Sinton, W. M. (1962). *In* "Physics and Astronomy of the Moon" (Z. Kopal, ed.), 1st ed. pp. 407–428. Academic Press, New York.
Six, N. F., Montgomery, C. G., Saari, J. M., and Shorthill, R. W. (1967). *AIAA Thermophys. Specialist Conf.*, 1967 Pap. No. 67-291.
Smith, B. G. (1967). *J. Geophys. Res.* **72**, 4059.
Thompson, T. W., and Dyce, R. B. (1966). *J. Geophys. Res.* **71**, 4843.
Ulrichs, J., and Campbell, M. J. (1969). *Icarus* **11**, 180.
Viskanta, R. (1965). *J. Heat Transfer* **83**, 143.
Viskanta, R., and Grosh, R. J. (1964). *Appl. Mech. Rev.* **17**, 91.
Watson, K. (1964). Ph.D. Thesis, California Institute of Technology (unpublished).
Wechsler, A. E., and Glaser, P. E. (1965). *Icarus* **4**, 335.
Wechsler, A. E., and Simon, I. (1966). "Thermal Conductivity and Dielectric Constant of Silicate Materials," Report, under contract NAS 8-20076 with Arthur D. Little, Inc., Cambridge, Massachusetts.
Wesselink, A. J. (1948). *Bull. Astron. Inst. Neth.* **10**, 356.
Wildey, R. L. (1967). *J. Geophys. Res.* **72**, 4765.
Wildey, R. L., Murray, B. C., and Westphal, J. A. (1967). *J. Geophys. Res.* **72**, 3743.
Winter, D. F. (1966a). *Int. J. Heat Mass Transfer* **9**, 527.
Winter, D. F. (1966b). *Icarus* **5**, 551.
Winter, D. F. (1967). *Icarus* **6**, 229.
Winter, D. F. (1970). *Radio Sci.* **5**, 229.
Winter, D. F., and Krupp, J. A. (1971). *The Moon* **2**, 279.
Winter, D. F., and Saari, J. M. (1969). *Astrophys. J.* **156**, 1135.
Winter, D. F., Bastin, J. A., and Allen, D. A. (1972). *In* "Lunar Thermal Characteristics" (J. W. Lucas, ed.). AIAA Press, New York.

Author Index

Numbers in italics refer to the pages on which the complete references are listed.

Subject Index

S

T

U

V

Y